JOURNAL OF EMERGING COMPUTER TECHNOLOGIES (JECT)

VOLUME: 1 ISSUE NO: 1 YEAR: 2021 PERIOD: JANUARY-JUNE

e-ISSN: 2757-8267

INTERNATIONAL, OPEN ACCESS, PEER-REVIEWED JOURNAL

JOURNAL OF EMERGING COMPUTER TECHNOLOGIES (JECT)

E-ISSN: 2757-8267

Volume :1 Number:1 Year:2021

Indexing and Abstracting

Index Copernicus

Academia.edu

Google Scholer

Academic Resource Index (Researchbib)

JOURNAL OF EMERGING COMPUTER TECHNOLOGIES

Volume: 1 **No:** 1 **Year:** 2021 **Period:** January-June

Head Office : İzmir

Type of Publication : It is published electronically twice a year, in June and December.

Journal of Emerging Computer Technologies

Address : Yaylacık Mh. 188 sk. No:4 D:3 Buca / Izmir, TURKEY
Web : http://ject.izmiracademy.com
Email : journalect@gmail.com
Email : ject@izmiracademy.com

Journal of Emerging Computer Technologies publishes scientific/original research articles. It is published electronically twice a year, in June and December. It is an **international scientific refereed journal** that publishes articles written in **English** and **Turkish**, and includes acedemicians from different countries in its boards. Uses "Double-Blind Peer Review" in reviewing processes. It has adopted the open access principle. No fee is requested from the authors for open access, processing, publication fee or otherwise. It is totally **FREE.**

The responsibility of the articles published in this journal in terms of science and language belongs to the authors. The articles published in the journal cannot be used without reference.

Policies and Conditions

Publication Ethics

* The publishing processes of the Journal of Emerging Computer Technologies (JECT) are based on the unbiased production, development and sharing of information using scientific methods.
Refereed articles are studies that ensure the application of the scientific method and impartiality. In the realization of scientific production, all components of the publication process; publishers, editors, authors, referees and readers must comply with ethical principles.
* In this context, the publication ethics and open access policy of Journal of Emerging Computer Technologies (JECT), in line with the guidelines and policies published by the Committee on Publication Ethics (COPE) in open access (For example, the "Publication Ethics Committee (COPE) Code of Conduct and Journal Editors" The Best Practice Guidelines for “Code of Conduct and Best Practice Guidelines for Journal Editors” and “COPE Best Practice Guidelines for Journal Editors”) require that all components of the publication process comply with ethical principles.
* In situations such as manipulating and distorting the data used in the articles and using fake data are detected, this situation will be officially reported to the institution where the article author works and the article will be rejected. Our journal has the right to request the output files of the analysis results from the authors according to the feedback given by the editor and/or referees.

Publisher's Ethical Responsibilities

* The publisher of the Journal of Emerging Computer Technologies (JECT), is a non-profit person who takes care of the public interest.

* Journal of Emerging Computer Technologies (JECT) has been serving as an important journal in the field of social sciences under the management of the editorial board or editorial board since 2019.

* Members of the Editorial Board of Journal of Emerging Computer Technologies (JECT) are assigned for three years, taking into account the branches of social sciences. As the head of the Editorial Board, an editor and, depending on the situation, one or more assistant editors and field editors can be determined. A new member is appointed to replace the leaving member. The term of the expired member can be extended. According to the subject of the article, if necessary, the editor may send an article to experts who are not members of the Board of Editors (at least doctor's title) for review as a field editor.

Responsibilities of the Editorial Board of Journal of Emerging Computer Technologies (JECT)

* The Editorial Board of Journal of Emerging Computer Technologies (JECT) is responsible for every article submitted to the Journal of Emerging Computer Technologies (JECT) and even all processes after its publication, and the Editorial Board knows this responsibility. This responsibility, considering only the public interest in decisions made on issues related to the journal; it requires making decisions independently without considering personal gain. The relationship between the Publisher and the Editorial Board is based on the principle of independence, and all decisions taken by the editors are independent of the publisher and other persons and organizations.
* The Editorial Board of Journal of Emerging Computer Technologies (JECT) strives to continuously improve the journal and increase its publication quality.

* The Editorial Board of Journal of Emerging Computer Technologies (JECT) ensures that journal policies such as publication, blind refereeing, evaluation process and ethical principles are determined and implemented.

* The Editorial Board of Journal of Emerging Computer Technologies (JECT) protects the copyright of the article authors published in Journal of Emerging Computer Technologies (JECT).

* The Editorial Board of Journal of Emerging Computer Technologies (JECT) is responsible for taking measures regarding intellectual property rights, unscientific-unethical behaviours, plagiarism, sending (citation) gangs during the publishing process of articles and journals.

* The Editorial Board of Journal of Emerging Computer Technologies (JECT) prepares an "Author's Guide" containing the information requirements of the author (s) and a "Referee Guide" containing the information that referees will need during the evaluation phase, and updates it when necessary.

* The Editorial Board of Journal of Emerging Computer Technologies (JECT) determines policies to motivate the referees and the author (s).

* The Editorial Board of Journal of Emerging Computer Technologies (JECT) keeps the records of each article and correspondence related to the journal electronically or in print.

Ethical Responsibilities of Editor, Assistant Editors and Field Editors

* Editors strive to meet the information needs of referees, author (s), researchers, practitioners and readers, to provide feedback when necessary, to act according to the principles of clarity in matters that require correction and explanation during the publication process.
* While deciding on the publication of articles, editors take care that the articles are original and contribute to the scientific literature, readers, researchers and practitioners.
* While making positive or negative decisions about articles, editors take into account the original value of the articles, their contribution to the field, the validity and reliability of the research method, the clarity of the expression, and the journal's purpose and scope.
* Editors take the submitted articles to the pre-evaluation stage as long as they have no significant problems, consider positive referee suggestions, and do not change the decisions made by the previous editor (s) unless there is a serious problem.
* Editors implement the policies of blind review and evaluation process, which are among the publishing policies of the journal, keep the identity information of the referees confidential, and ensure that each article is evaluated objectively and in time.
* Editors send articles by taking into account the field of expertise of the editors and referees, and they support the impartial and independent evaluation.
* Editors consider whether there is a conflict of interest between editors, referees and author (s) for unbiased evaluation of the article.
* Editors seek to ensure that the referee pool consists of a wide variety and is constantly updated.
* Editors avoid unscientific reviews that do not follow academic etiquette.
* Editors ensure that the journal publishing processes are operated in accordance with publishing policies and guidelines, inform their field of duty in the process about the developments in publication policies, and prepare training programs when necessary.
* Editors are ineffective communication with everyone involved in the publishing process and hold meetings at regular intervals.
* Editors ensure the protection of personal data in reviewed articles; protect individual data of authors, referees and readers.
* Editors; They pay attention to the protection of human and animal rights in the articles, they care about documenting the explicit consent of the participants of the article, the approval of the ethics committee of the participants of the article, and they reject the article when they do not have permission for experimental research.
* Editors; they take precautions against misconduct. When there are complaints about misconduct, he makes an objective investigation and shares the findings on the subject.
* Editors ensure that errors, inconsistencies or misdirections in the articles are corrected.
* Editors protect the intellectual property rights of the published articles and defend the rights of the journal and the author (s) in case of violation. In addition, they take necessary measures to ensure that the content of published articles does not infringe the intellectual property rights of other publications; They do the authenticity-similarity check.
* Editors take into account the consistent criticism of the articles published in the Journal and give the right to reply to the author (s) of the criticized articles.
* Editors also consider studies with negative results.
* Editors examine the complaints submitted to the Journal and make the necessary explanations.

Ethical Responsibilities of Referees

* In the Journal of Emerging Computer Technologies (JECT) article evaluation process, the two-way blind refereeing principle is applied, in which the referees of the author (s) and the referees do not recognize the authors. The article evaluation forms and the notes stated on the text and the correction requests are forwarded to the author (s) by the editors through the journal management system. Referees who will evaluate the articles submitted to Journal of Emerging Computer Technologies (JECT) should bear the following ethical responsibilities:

* Referees should only agree to evaluate articles related to their field of expertise.
* Referees should make the assessment in neutral and confidentiality. In accordance with this principle, they should destroy the articles they have reviewed after the evaluation process, but use them after publication. Nationality, gender, religious belief, political belief and commercial concerns should not disturb the neutrality of the evaluation.
* When the referees understand that there is a conflict of interest, they should refuse to evaluate the article and inform the editors.
* Referees should make the evaluation in a constructive language in accordance with academic etiquette; Avoid personal comments containing insults and hostility.
* The referees should evaluate the article, which they accepted, within the time limit.

Ethical Responsibilities of Authors

Ethical responsibilities of the author (s) who apply for an article to Journal of Emerging Computer Technologies (JECT):
* Author (s) should not send an article that they have published elsewhere or sent to be published, or more than one article at the same time to Journal of Emerging Computer Technologies (JECT).
* Author (s) should submit an original article to Journal of Emerging Computer Technologies (JECT).
* The author (s) should correctly refer (cite) to the sources they use during the writing of the article in line with ethical principles.
* The names of persons who do not contribute to the article should not be written as authors, and it should not be suggested to change the order of the author of an article submitted for publication, to remove or add authors.
* Individuals who have a conflict of interest and a common interest related to the article for publication must be notified to editors.
* If information or raw data is requested from the author (s) during the evaluation process, they should submit the expected information to the Editors.
* The author (s) must document that the rights to use the data they use in their articles, permission for research-analysis or the consent of the participants on which they have researched.
* The author (s) should contact the editor to inform, correct or retract when they notice the error in the evaluation and early viewing or electronically published article.
* The author (s) have obtained the approval of the ethics committee for research that requires data collection by quantitative or qualitative methods such as experiments, questionnaires, scales, interviews, observations, focus group studies that require ethics committee decision; The name of the ethics committee, the date and number of the decision should be indicated on the first-last page of the candidate article and in the method section, and the document showing the ethics committee decision should be uploaded to the system with the application of the article. In addition, case reports should include information about obtaining informed consent / obtaining consent form in the article.
* The author (s) should present in the article the evidence that they adhere to ethical principles in the data collection process (such as obtaining permission from others to use documents such as scale, questionnaire, photograph). It should be stated in the articles that the copyright regulations for research and publication ethics and intellectual and artistic works are complied with. If the research was carried out on human and animal subjects, it should be reported that the research was carried out in accordance with international reports, guidelines, etc.
* Ethics committee approval is not required from the author (s) for review articles. However, in articles that do not require an ethics committee decision, it should be stated that an ethics committee decision is not required, on the first-last page and method section of the article.

Notifying the Editor of Non-Compliance with Ethical Principles

In the Journal of Emerging Computer Technologies (JECT), if an unethical situation regarding editors, referees, authors or an article that is in the process of evaluation, early appearance or an unethical situation is encountered, it should be reported to journalect@gmail.com.

Research Ethics

The following codes of ethics on research ethics contribute to research objectives, support values ??for collaboration, ensure researchers' responsibility to the public, provide credibility with public support, and reinforce various moral and social values:
1. The first rule is openness. In other words, the researcher has to share the data, materials, equipment, resources and findings he uses with the public. It should be open to criticism and new ideas.
2. Honesty and truthfulness: The researcher should not make up, steal, or misinterpret his findings. It must report results, methods, procedures accurately and not mislead colleagues, sponsors and the public. If he/she abuses research, it is also an unethical behaviour.
3. The researcher should not cause any harm to the object or subject of interest. Consequently, the researcher must respect people's privacy and secrets and should not use personal names unless permitted in his research. In addition, the researcher should not hide his identity.
4. Intellectual property: Patents, copyrights and other intellectual property must be respected. Unauthorized data, methods and results should not be used and contributors should be referenced.
5. Confidentiality: Confidential communications, personnel records, commercial and military secrets, patient records must be protected.
6. Being objective: biased experimental design, data analysis and interpretation should be avoided.
7. Integrity and consistency must be ensured.
8. Errors and omissions made due to carelessness should be avoided through a serious thorough examination of the work.
9. Social responsibility: Researches that may cause social damage should be avoided.
10. Discrimination: It should not be discriminated against by using gender, race, origin and other factors that are not related to scientific authority and integrity.
11. Relevant laws should be known and obeyed.
12. Caring for animals: Animals used during the research should be given due care and respect. Poorly designed and unnecessary animal experiments should not be done.

13. Protecting people: Risk and harm should be minimized in humanitarian research, human dignity, privacy and autonomy should be protected. Special precautions should be taken in research involving vulnerable persons with a child, developmental or cognitive disability, living in a nursing home, homeless or without legal status. The burden and benefits of research should be distributed fairly.
Planning and implementation of research on humans and animals, whose framework is drawn by national and international regulations and regulations, are the most important topics of research ethics. Research ethics is a concept that is mostly used for medical research, but undoubtedly, research conducted in social sciences is also within the scope of research ethics in a different dimension.

Obtaining Legal / Private Ethics Committee Permit

TRDIZIN EXPLANATION and DISCLOSURE of the Code of Ethics The articles related to the Code of Ethics in the criteria in the previous years were detailed in 2020 with explanations, assuming that permissions were obtained for studies requiring ethical committee approval in the field of research, and the issue of "including information about permission in the article" was added to the criteria.
Researches requiring Ethics Committee permission are as follows:
1) All kinds of research conducted with qualitative or quantitative approaches that require data collection from participants using a questionnaire, interview, focus group study, observation, experiment, interview techniques.
2) Use of humans and animals (including material/data) for experimental or other scientific purposes,
3) Clinical researches on humans,
4) Research on animals,
5) Retrospective studies in accordance with the law on the protection of personal data,
Also;
Stating that "Informed Consent Form" was obtained in case presentations,
Obtaining and indicating permission from the owners for the use of scales, questionnaires, photographs belonging to others,
Statement of compliance with copyright regulations for the intellectual and artistic works used
Retrospective ethics committee approval is not required for articles that have used research data before 2020, that have been produced from master's / doctoral studies (should be specified in the article), that have been published in the journal in the previous year, accepted but not yet published.
Researchers who are not members of the university can also apply to the Ethics Committees in their regions.
Also:
Whether an ethics committee approval and/or legal/special permission is required for articles to be published in journals should be stated in the article. If these permissions are required, it should be clearly presented from which institution, on what date and with which decision or issue number the permission was obtained.
If the study requires the use of human and animal subjects, the study should be included in an international declaration, guide, etc. It must be declared appropriate

Open Access Policy

Journal of Emerging Computer Technologies (JECT) has adopted the "open access policy" by considering the principle that increasing the global exchange of the knowledge generates useful results for humanity.Journal of Emerging Computer Technologies (JECT) is an "open access" journal the content of which is available to all users freely. Users can read the articles in full text, download, issue, print out and link to articles citing the journal without taking permission from the publisher or the author(s).
Our open-access policy is in line with the Budapest Open Access Initiative:
(http://www.budapestopenaccessinitiative.org/read).

https://creativecommons.org/licenses/by-nc/4.0/

CONTENT

Journal of Emerging Computer Technologies
Research Article
Received: *2020-10-04* | ***Reviewing:*** *2020-10-28* & *2020-11-18*| ***Accepted:*** *2020-12-02* | ***Online:*** *2020-12-10* | ***Issue Date:*** *2021-06-30*

Log Analysis with Hadoop MapReduce

Gligor Risteski
Faculty of Communication Network and Security
University of Information Science and Technology, Ohrid, Macedonia
gligorr@gmail.com

Mihiri Chathurika
Faculty of Communication Network and Security
University of Information Science and Technology, Ohrid, Macedonia
mihiritxla@gmail.com

Beyza Ali
Faculty of Communication Network and Security
University of Information Science and Technology, Ohrid, Macedonia
beyzaali34@gmail.com

Atanas Hristov
(Corresponding Author)
Faculty of Information and Communication Sciences
University of Information Science and Technology, Ohrid, Macedonia
atanas.hristov@uist.edu.mk,
ORCID: 0000-0003-2741-8370

***Abstract*— Pretty much every part of life now results in the generation of data. Logs are documentation of events or records of system activities and are created automatically through IT systems. Log data analysis is a process of making sense of these records. Log data often grows quickly and the conventional database solutions run short for dealing with a large volume of log files. Hadoop, having a wide area of applications for Big Data analysis, provides a solution for this problem. In this study, Hadoop was installed on two virtual machines. Log files generated by a Python script were analyzed in order to evaluate the system activities. The aim was to validate the importance of Hadoop in meeting the challenge of dealing with Big Data. The performed experiments show that analyzing logs with Hadoop MapReduce makes the data processing and detection of malfunctions and defects faster and simpler.**

Keywords*— *Hadoop, MapReduce, Big Data, log analysis, distributed file systems.

I. INTRODUCTION

The term "Big Data" is gaining more popularity every day. The first thing we should know about it is that it does not have a commonly held definition. Basically, as one can understand from its name, Big Data means a big amount of data. Sethy, R. [1] in his article defines "Big Data describes any massive volume of structured, semi-structured and unstructured data that are difficult to process using a traditional database system."

Researches show that data volumes are doubling every year. Although there is not a specific reason behind this rapid growth rate, the new data sources, contribute to that growth highly. Smartphones, tablet computers, sensors, and all other devices that can be connected to the internet generate a vast amount of data. Enterprises improve their technological infrastructures and adopt more powerful platforms, which play an important role in the growth rate of the data that is generated [2].

II. HADOOP

Hadoop is a collection of open-source utilities which allows the use of network to deal with the problems which include big amounts of data. Hadoop provides framework for distributed storage and framework for processing big data with MapReduce programing model. The core of Apache Hadoop framework contains the following parts:

- Hadoop Common
- Hadoop Distributed File System (HDFS)
- Hadoop YARN
- Hadoop MapReduce

Hadoop common represent libraries and utilities which are needed by other Hadoop parts in order to operate. Hadoop distributed File System (HDSF) is a distributed file-system that stores data on commodity hardware, allowing very high bandwidth across the cluster. Hadoop YARN is a platform responsible for managing computing resources in clusters and uses them for scheduling users' applications. Lastly, Hadoop MapReduce is a programing model for big data processing in the cluster.

Java is a programming language that is mostly used for writing MapReduce programs but Hadoop allows the use of any programming language to write MapReduce programs.

To exploit the parallel processing that Hadoop gives, we have to express our query as a MapReduce job. After some local, little scope testing, we can have the option to run it on a cluster of machines.

The Hadoop structure is demonstrated in Figure 1.

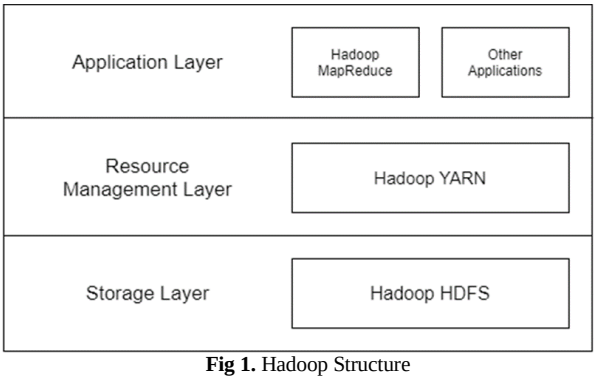


Fig 1. Hadoop Structure

A. Hadoop Distributed File System (HDFS)

Hadoop Distributed File system is one of the most reliable storage systems designed to store a smaller number of large files rather than a greater number of small files. Among many of the features HDFS provides, the fault-tolerant storage layer can be mentioned as one of the most important features.

Replication of data in the Hadoop file system helps the user to attain this feature. Even in situations where hardware failure happens, the data reliability is still high [3].

Apache Yarn, introduced in Hadoop 2.x, is the resource management layer of Hadoop, which is also used for job scheduling and data operating system. It allows different data processing engines to run and process data stored in HDFS. Some of these processing engines are graph processing, interactive processing, stream processing, and batch processing. Data processing platform Yarn has the functionality named MapReduce, which empowers Hadoop by allowing processing numerous different frameworks on the same hardware where Hadoop is deployed [4].

HDFS was developed using distributed file system design and is designed using low-cost hardware. It is more fault-tolerant than other distributed systems. HDFS is capable of holding larger amounts of data providing easy access and parallel processing. In HDFS files are stored across multiple machines in order to prevent possible data losses in case of system failure [3]. Features of HDFS are: distributed storage and processing, command interface, checking the status of clusters, streaming access to the file system, file permission, and authentication.

- Distributed storage and processing
- Command interface to interact with HDFS
- Checking the status of cluster easily
- Streaming access to file system data
- File permissions and authentication

1) HDFS Architecture

HDFS follows the master-slave architecture. It has elements such as namenode, datanode, and blocks, where the built-in servers of the first two elements enable users to easily check the status of the cluster. Similarly, they are commodity hardware [point]. In Figure 2 the architecture of the Hadoop file system is demonstrated.

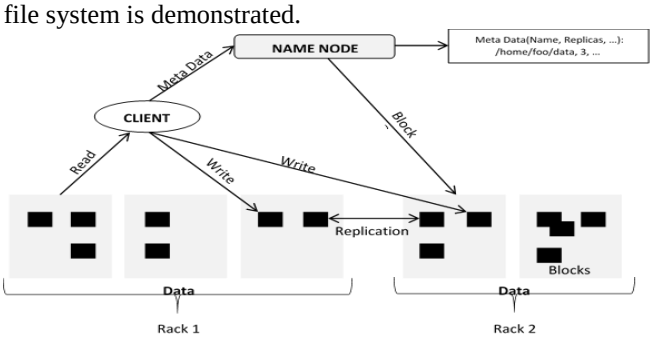


Fig 2. Architecture of a Hadoop File System

- Namenode

The namenode, containing GNU/Linux operating system and the namenode software, can act as the master server running on commodity hardware. Renaming and opening/closing files/directories are executed in this element. It also manages the file system and regulates clients' access to files.

- Datanode

Similar to namenode, the datanode too, contains the GNU/Linux operating system and datanode software. For every node in a cluster, there exists a datanode, which is able to manage the data storage of the system. Depending on the instructions of the namenode, datanode performs block creation, block deletion, and block replication. It also performs read/write operations on the file system when requested by a client.

- Block

. Usually, the user data are stored in files such that a file in a file system is divided into segments and then are stored in individual data nodes. Those file segments are called blocks. A block is the minimum amount of data that HDFS can read/write. The block size can be increased if needed but the normal size is 64MB.

B. Apache Hadoop Yarn

To create a split between Resource Manager (RM) and Application Master (AM), YARN separates the functions of resource management and job scheduling into separate daemons. An application can be an individual job or a DAG of jobs.

- ResourceManager (RM)

Resource Manager (RM) together with the Node Manager (NM) comprise the data-computation framework. The RM adjudicates the resources in the system and NM is responsible for containers and monitoring resource usage. In other words, RM is the ultimate authority and NM is the framework agent. RM has two main components known as Scheduler and Application Manager.

- Scheduler

Scheduler controls the allocation of resources to the several running applications. It performs the scheduling function depending on the resource requirements of applications, but it does not take any responsibility for performing monitoring, tracking the status of applications or application or hardware failures.

- ApplicationsManager

Applications Manager: responsible for restarting the AM container in case of a failure. It also accepts job submissions.

- ApplicationMaster (AM)

The AM is responsible for requesting resources from the RM and then executing and monitoring the tasks. It works together with NM when executing the tasks.

The main idea behind using YARN on Hadoop is the notion of resource reservation via the Reservation System. The Reservation System tracks the resources, performs admission control for reservations, and reserves resources to ensure the execution of important jobs [4].

YARN supports the notion of Federation via the YARN Federation feature. The idea behind this is to scale Yarn up to very large amounts of nodes by wiring YARN clusters and sub-clusters. The Federation feature makes this transparent wiring of clusters appear as a single big cluster [5].

III. The Role of MapReduce In Handling Big Data

Big Data means "big power" when handled efficiently. It can give new aspects to the enterprises, like which strategy will increase the profitability, which customers buy which products, the current situation of the company versus the situation of the competitors, and so on.

As the data comes from different sources and different structures it is important to categorize it with respect to some characteristics of the data. The most important and the most known characteristics of Big Data are known as the "3Vs of

Big Data" where the Vs stands for volume, variety, and velocity. Volume refers to the amount of data, variety refers to the type of data, i.e. text, image, video, etc. and velocity refers to the speed at which the data comes from different sources [6].

Big Data comes with its own set of problems that need to be resolved. Processing power, storage, data issues, and cost are the most important problems. The old techniques for working with or analyzing information are not enough to deal with Big Data. Therefore, new technologies are needed and this is where MapReduce comes into the picture [2].

A. MapReduce

MapReduce is a programming model that is used for accessing and processing big data stored in HDFS. Programs written in MapReduce are executed on a distributed system where big data is split into smaller chunks of data and are executed in parallel [7]. MapReduce has two functionalities, Map () and Reduce (). This model has been used in Google's search index, machine learning, and statistical analysis [8]. Implementation of MapReduce is highly scalable and easy to use. The run-time system allows programmers with no prior knowledge or experience with parallel processing to utilize the resources of distributed systems easily, by handling details like partitioning the inputs, scheduling the program's execution, handling failures, and managing inter-machine communication [7, 9].

Although it is impossible to prevent failures, the objective is to minimize the probability of failure to a level that will not harm the overall process. Two methods that would help to increase the "fault tolerance" in Big Data are the following [6]:

- First divide the whole computation into smaller tasks and then assign each task to a different node.
- Assign a node to observe if the other nodes are working properly. In case a node fails to complete its task, the task is restarted. But this may cause a complication in the process if some tasks are recursive.

MapReduce is one of the core building blocks of processing in the Hadoop framework. Hadoop uses the MapReduce algorithm to run the applications in parallel. It provides the necessary solution to keep the process going since it can survive failures without losing data. In 2004 Google published about MapReduce technology [7]. MapReduce comprises of two distinct tasks: Map and Reduce. Mapping is the first phase and Reducing happens after the Mapping phase is completed. In Map phase data is processed and key-value pairs are produced. This is known as the map job. Then the produced key-value pairs are fed into the Reducer. After collecting all the key-value pairs from all of the map jobs the Reducer groups the pairs into a smaller set of key-value pairs, producing the final output [7, 8].

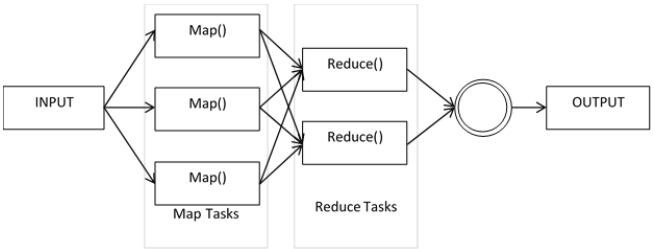


Fig 3. How does MapReduce work?

B. Algorithm of MapReduce

MapReduce program has three stages of executing: map, shuffle, and reduce.

- Map stage: at this stage, the map job processes the input data which is in the form of a file or directory. As an output of the map job several small chunks of data are produced.
- Shuffle stage: at this stage, the output of the map stage is accepted and the relevant records are consolidated.
- Reduce stage: reduce stage acts together with the shuffle stage. The reducer processes the data that comes from the map stage and produces a new output which is then stored in the Hadoop file system.

During the process, a task from Map and Reduce phases are sent to the appropriate servers in the cluster. All the data passing details like issuing tasks, verifying task completion, copying data between nodes, are managed by the MapReduce framework. Because the computing is performed on nodes using the data from local disks the network traffic is reduced significantly. After the tasks are completed the results are sent back to the Hadoop server [10].

C. MapReduce with Python

'mrjob' is a Python library for MapReduce for writing and running Hadoop streaming jobs. It is created by Yelp. When a MapReduce job is written using 'mrjob', it can be tested locally and run on a Hadoop cluster or in the cloud. Using 'mrjob' for writing MapReduce applications has many advantages. Some of them are:

- It is a dynamically developing framework.
- It has extensive documentation.
- Installing Hadoop is not enforced. Applications written using 'mrjob' can be executed and tested without installing Hadoop.
- It allows the MapReduce applications to be written in a single class rather than writing separate programs for Map and Reduce phases.

Although it provides a great solution, 'mrjob' has its disadvantages. The most important disadvantage is that it does not provide the level of access to Hadoop that other APIs provide. This is because it is a simplified library [11].

IV. LOG ANALYZER WITH HADOOP

Log analysis is both art and science which aims to make sense out of computer-generated records. These records are called log or audit trail records. The process of creating these records is known as data logging [12]. Some of the most important reasons for performing log analysis are:

Understanding user behavior

- System troubleshooting
- Proper resource allocation
- Improved business operations
- Improved security
- Achieve compliance

Data centers generate thousands of terabytes or petabytes of log files every single day. It is very challenging to store and analyze these data not only because of its large volume but also because of the different structure of log files. Due to not being able to deal with a large volume of log files efficiently, conventional database solutions run short for the needs in log analysis. As a result of the comparison of SQL DBMS and Hadoop MapReduce in [13], Hadoop MapReduce overperforms DBMS in the means of tuning up with the task and loading data. As it can be seen from this result, with the unprecedented increase in the data generated traditional methods fall short with providing a solution for data analysis. This is, exactly, the point where the new technologies stepped in [8]. Hadoop MapReduce has a wide area of applications for Big Data analysis [3],[9],[11]. The true power of Hadoop lies in its ability to scale up to a great number of computers, where each computer has several processor cores, by connecting commodity computers to work in parallel. This plays an important role in log analysis as it can benefit thousands of nodes which will store multiple blocks of log files.

In this paper, we propose an idea on how Hadoop can be used to analyze web server logs, in our case Nginx access log. Web server access logs are generated by the web servers all the time, recording all accesses on the hosted web pages. This means that the access logs can be very big. The web access log contains information about time, IP addresses, browser type, etc. All of this information is important for the system administrators as it provides information about system usage, security, and system troubleshooting.

This idea is proposed for analyzing one kind of logs only, but Hadoop can be used in every situation where big log files are generated, such as system logs, logs from some business application, etc.

The real-world usage (practical usage) of this system can be implementing it as a base of a larger system used by many users for log analyzing.

A. *Environment Setup*

We have installed Hadoop for demonstrating purpose on two virtual machines hosted on Digitalocean. Each virtual machine has 2 cores CPU and 8GB of RAM memory (installing is explained in details in Installing Hadoop section) with installed Ubuntu 16.04. Also, for testing purpose we have installed Hadoop on one local virtual machine with 1 core CPU and 7GB of RAM memory on hosted hypervisor VMware Workstation Pro 15 using Bitnami Hadoop Stack image. In this project the log generated from Python Fake Logs script is used [14].

V. RESULTS

Our aim was to validate the role and importance of Hadoop in meeting the challenge of dealing with Big Data, by an analysis of log files in order to evaluate the system activities. We benefitted from Hadoop's ability to scale up to as many computers as needed. Firstly, log files are broken up into blocks with MapReduce class created with mrjob python library (dividing into blocks is explained in section 3.1). Blocks are then distributed over the nodes in a Hadoop cluster. With parallel computation, the job is divided into a number of tasks, which in return improves the performance. In our log analysis, we used Python's 'mrjob' library.

We use a dummy log file only for testing. Our log file is the Nginx access log file. The log file contains information about:

- visitor's IP address,
- date and time of the access,
- visited page,
- type of request,
- type of the user's web browser,
- type of the user's operating system.

In our case, two reports are extracted from the Nginx access file as an example. These two reports are only an example of what can be achieved. The following charts demonstrate the results of our experiments. In Figure 4 the number of visits is shown based on the hours the visitor visited the websites. We can see that the number of visits is the highest at 20:00 – 21:00. The number of visits is the lowest at 10:00 – 11:00.

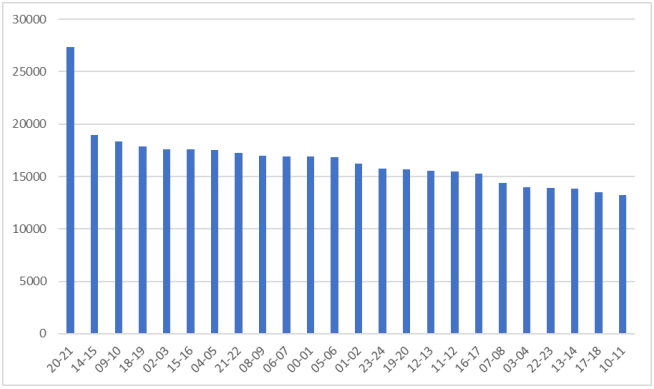


Fig 4. Number of visits by hour

The second report analyzes which are the most visited pages. It is possible to make many other reports depending on the user's needs. For example, the most used browser, most used operating system, detecting suspicious behavior, etc.

In Figure 5 the distribution of the visits is displayed with regard to the most visited 10 sites.

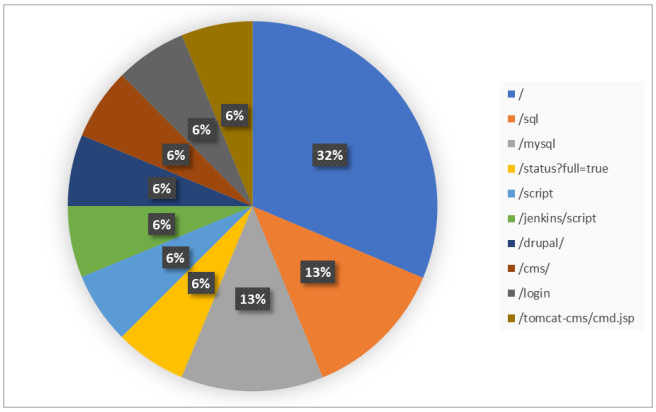


Fig 5. Top 10 most visited sites

VI. DISCUSSION

Log data grows very quickly as they are generated at a record rate. Processing power, storage, and cost are some of the biggest challenges that come with this Big Data. Using conventional methods to deal with such data is too costly in terms of both time and money. Hadoop provides frameworks

for distributed storage and processing for Big Data, which makes it possible to use a network for processing and analyzing big amounts of data. In this way, companies save time and money when trying to diagnose problems, solve issues, or obtain the knowledge they wouldn't be able to obtain otherwise. The ability to store as well as distribute large datasets across numerous servers in a cost-effective way makes Hadoop very advantageous with regard to the traditional relational database management systems. Moreover, Hadoop MapReduce processes terabytes of data in minutes and by executing tasks in parallel shortens the processing of data by a considerable amount.

VII. CONCLUSION

Log analysis is important for businesses and system administrators in many aspects, as they give information about malfunctions, defects, security-related issues, and so on. On the other hand, being generated constantly provides a good example of Big Data, therefore a good example of our experiments. Our analysis shows that analyzing logs with Hadoop MapReduce makes the detection of malfunctions, defects, and so on faster and simpler. The results also show that there is not a big difference if we run the code in a real cluster or in the test environment.

REFERENCES

[1] Sethy, R. et al. Big Data Analysis using Hadoop: A Survey. International Journal of Advanced Research in Computer Science and Software Engineering 5(7), 2015, pp. 1153-1157.

[2] Schneider, R.D. Hadoop For Dummies, Special Edition. John Wiley & Sons Canada, Ltd. 2012.

[3] Borthakur, D. HDFS architecture. Document on Hadoop Wiki. http://hadoop. apache. org/common/docs/r0 20. 2010.

[4] Vavilapalli, V. K.; et al. Apache hadoop yarn: Yet another resource negotiator. Proceedings of the 4th annual Symposium on Cloud Computing. ACM, 2013.

[5] Hadoop, Apache. Hadoop Archives Guide. The Apache Software Foundation, http:// hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html (2019). Retrieved Oct. 15, 2019.

[6] Kaur, I. et al. Research Paper on Big Data and Hadoop. IJCST, 7(4), 2016, pp. 50-53.

[7] Dean, J. and Ghemawat, S. MapReduce: Simplified data processing on large clusters. Proceedings of Operating Systems Design and Implementation, 2004.

[8] Yang, H. et al. Map-reduce-merge: simplified relational data processing on large clusters. Proceedings of the ACM SIGMOD international conference on Management of data. ACM, 2007.

[9] Rohloff, K. and Schantz, R.E. High-performance, massively scalable distributed systems using the MapReduce software framework: the SHARD triple-store. Programming support innovations for emerging distributed applications. ACM, 2010.

[10] Point, Tutorials. Retrieved Oct. 15, 2019 from Internet Site https://www.tutorialspoint.com.html. Tutorials Point.

[11] Miner, D. and Radtka, Z. Hadoop with Python. O'Rilley Media. 2016.

[12] Log analysis https://en.wikipedia.org/wiki/Log_analysis

[13] Sayalee Narkhede and Tripti Baraskar - Hmr Log Analyzer: Analyze Web Application Logs Over Hadoop MapReduce International Journal of UbiComp (IJU), Vol.4, No.3, July 2013.

[14] "Python Fake Logs." Internet: https://github.com/s4tori/fake"Python Fake Logs." Internet: https://github.com/s4tori/fake-logs.-logs.

Journal of Emerging Computer Technologies
Research Article
Received: *2021-01-26* | ***Reviewing:*** *2021-01-26* & *2021-02-07* | ***Accepted:*** *2021-02-07* | ***Online:*** *2021-02-07* | ***Issue Date:*** *2021-06-30*

Vehicle Detection with HOG and Linear SVM

Nikola Tomikj
Ss. Cyril and Methodius University
Faculty of Computer Science and Engineering
1000 Skopje, North Macedonia
tomikj.nikola@students.finki.ukim.mk

Andrea Kulakov
Ss. Cyril and Methodius University
Faculty of Computer Science and Engineering
1000 Skopje, North Macedonia
andrea.kulakov@finki.ukim.mk

***Abstract*—In this paper, we present a vehicle detection system by employing Histogram of Oriented Gradients (HOG) for feature extraction and linear SVM for classification. We study the influence of the color space on the performance of the detector, concluding that decorrelated and perceptual color spaces give the best results. An in-depth analysis is carried out on the effects of the HOG and SVM parameters, the threshold for the distance between features and the SVM classifying plane, and the non-maximum suppression (NMS) threshold on the performance of the detector, and we propose values that illustrate good performance for vehicle detection on images. We also discuss the issues of the approach and the reasons for its mediocre performance on videos. Finally, we address these issues by presenting ideas that can be considered for improving the system.**

***Index Terms*—computer vision, machine learning, HOG, SVM, color space, vehicle detection, autonomous vehicles**

I. Introduction

In the past few years, autonomous driving has been gaining a lot of interest and it is expected to be the next big thing in the automotive industry. One of the main challenges in the development of the intelligence that powers the autonomous vehicles is its ability to detect obstacles like pedestrians, other vehicles and objects on the road. This ability provides the safety required to make autonomous vehicles mainstream. Different techniques for preceding vehicle detection have been developed throughout the literature, from traditional computer vision techniques to deep learning ones.

Dalal and Triggs [1] describe the HOG method in their breakthrough paper. They use the method for human detection. They performed thorough experiments with the parameters of the method. The values for the parameters that they found to be optimal for human detection are akin to the values that we found to be optimal for vehicle detection.

Creusen, Wijnhoven, Herbschleb, *et al.* [2] experimented with different color spaces and concluded that the choice of a color space significantly influences the performance of the HOG detector, and that the optimal color space choice depends on the type of object that the detector is trying to detect. They showed that for decorrelated color spaces like HSV, the performance when the H-channel is used as a single channel detector is almost identical to the performance in the HSV space. They concluded that this indicates that saturation and intensity information is largely irrelevant, and that color is the dominant feature. They also concluded that HSV and RGB are less suitable for traffic sign detection and detection of objects that have large color variation in general. They found out that the LAB and YCrCb color spaces provide the best performance, and that is probably due to the availability of two dedicated color channels.

Mao, Xie, Huang, *et al.* [3] took very similar approach to the one described in this paper, using HOG and linear SVM as well, and they developed effective and robust preceding vehicle detection system that can achieve high reliability target detection and low false positive rate.

An interesting and very useful approach was proposed by Arróspide, Salgado, and Camplani [4]. They overcame the computational limitations of the standard HOG, which yields excellent performance but can hardly be used in a real-time environment. They developed alternative HOG descriptors which are designed to be cost effective by making use of the previous knowledge on vehicle appearance.

Lately, deep learning approaches have been gaining popularity and several authors have experimented and produced state-of-the-art results. Fast R-CNNs are used extensively for object detection as proposed by Girshick [5]. Another important mention is YOLO introduced by Redmon, Divvala, Girshick, *et al.* [6].

In this paper we develop a pipeline for detecting preceding vehicles using HOG and linear SVM. Our aim is to study the HOG and linear SVM applicability and potential for vehicle detection. We experiment with the HOG and SVM parameters, the threshold for the distance between features and the SVM classifying plane, and the NMS threshold, to examine their influence on the performance of the system and to reveal values that provide good performance.

II. Methodology

We implemented a vehicle detection pipeline in Python that detects vehicles in images and videos recorded with a dashcam, using HOG and Linear SVM. The pipeline is developed in Python 2.7 [7] and OpenCV 3.4.3 [8].

The details of each component in the pipeline are presented as follow:

A. Data Preprocessing

The labeled data come from a combination of the GTI vehicle image database [9] and KITTI vision benchmark suite [10]. The data are png color images with dimensions 64×64,

which is convenient for computing the HOG descriptors of the images. The labeled data contain 8792 vehicle images and 8968 non-vehicle images. The non-vehicle data also contain some images extracted from a real dashcam video with hard negative mining to reduce the number of false positives.

Because all training images have the same dimensions, their HOG feature vectors have the same length and can be used to train an SVM. So, preprocessing is not necessary, however converting the images to certain color spaces could potentially increase the performance of the detector.

B. Feature Extraction

After the preprocessing stage, feature extraction is performed by computing the HOG descriptors of every preprocessed image from the labeled dataset. These descriptors are used to train and test a linear SVM.

C. Training a Linear SVM

This is 2-class classification problem, so the viable SVM type options that OpenCV offers are C_SVC and NU_SVC types. As they are very similar, we decided on using NU_SVC.

We were bound to use linear SVM because only the primal form of a linear SVM can be passed as an argument to the HOGDescriptor struct setSVMDetector function. This function allows the usage of the performant detectMultiScale function which this pipeline is trying to make use of.

D. Vehicle Detection

The last step is to perform vehicle detection on real dashcam data. For this purpose, we used the detectMultiScale function to detect vehicles in images and videos. This function performs a sliding windows search using windows with different sizes, so that vehicles with arbitrary dimensions can be detected in the input image or video.

III. Results

A. Color Space

To determine the color space that provides the best performance, experiments using various color spaces were performed, using constant values for the HOG and SVM parameters.

Table I gives comparison of the performance for 8 different color spaces based on the error as a percent of misclassified images from the test set when the images are in certain color space. The values of the HOG and SVM parameters used for the experiments are given in table II. These are some common default values for these parameters.

Oddly enough, BGR gives the best performance, or the least percentage of misclassified samples. This was unexpected because the BGR color space has strongly correlated channels and is non-perceptual. But, although it achieves the best performance during the testing phase, it performs poorly when detection is performed on real images and videos. We think that the reason is that real images and videos have a lot of background color variation, so using BGR which has strongly correlated channels will cause many misdetections. That is the case because HOG calculates separate gradients for each color channel and the one with the largest norm is taken as the pixel's gradient vector. LAB, LUV, YCrCb and YUV all achieve similar performance, and follow after BGR. Any of these 4 color spaces is a viable option because the difference between their performance is very small. We decided to use the YUV color space because it has the smallest error percentage in table I and it works well with our test data.

Table I
Color Space Performance

Color Space	Error
BGR	2.5732%
GRAY	3.5867%
LAB	3.3531%
LUV	3.2573%
HLS	5.7742%
HSV	6.0923%
YCrCb	3.2404%
YUV	3.1757%

Table II
Parameters Testing Values

Parameter	Value
Detection window size	64×64
Block size	16×16
Block stride	8×8
Cell size	8×8
Bins	9
Discrete derivative mask size	1
Gaussian smoothing window parameter	-1 (means no smoothing)
Block normalization type	L2-Hys
L2-Hys threshold	0.2
Gamma correction	False
Signed gradient	False
Nu	0.09

B. HOG Parameters

HOG is the key component in the pipeline, so the choice of the values for the HOG parameters is a very important one. We experimented with different values for these parameters and found out that the ones suggested by Dalal and Triggs [1], shown in table II are performing reasonably well. The specific effects that the values of these parameters have on the performance i.e. the linear SVM testing error, are the following:

- Because the dataset images are 64×64 pixels, it is reasonable to use a 64×64 detection window size. Decreasing or increasing this size will decrease the performance because the images will contain less information, or it will make the descriptor more sensitive to noise, respectively.
- Block size, block stride and cell size depend on each other's value and influence the performance together, so they must be tested together. Table III shows the results of the experiments and gives insight about the effects of these three parameters on the performance. The interpretation of these results is that overlap of $3/4$ slightly improves the performance, however it drastically

Table III
HOG PARAMETERS PERFORMANCE

Block Size	Block Stride	Cell Size	Error
8×8	4×4	4×4	3.2751%
8×8	8×8	8×8	6.5991%
16×16	4×4	4×4	2.3341%
16×16	8×8	8×8	2.6858%
16×16	16×16	16×16	7.7984%
32×32	8×8	8×8	2.3649%
32×32	16×16	16×16	2.7703%
32×32	32×32	32×32	16.4527%

increases the time required for extracting the HOG features and training the SVM which makes it not worth it for such a small performance increase. $1/2$ overlap is enough.

- Discrete derivative masks of size 1 without Gaussian smoothing work best. Using larger masks decreases performance and smoothing decreases performance significantly. However, larger masks and Gaussian smoothing can be useful on real videos.
- Using more than 9 bins with unsigned gradients will not improve the performance significantly. Decreasing the number of bins decreases the performance. So, it is appropriate to use 9 bins for vehicle detection.
- The only block normalization type currently available in OpenCV is L2-Hys. The optimal L2-Hys threshold is 0.2, increasing or decreasing this threshold decreases the performance.
- Using gamma correction increases the error for around 0.5%, however it can be useful on real videos.
- Dalal and Triggs [1] used unsigned gradients for human detection in their paper, but they suggested that for some other tasks like vehicle detection sign information helps substantially. However, we found out that this is not the case. Signed gradients seem to cause overfitting of the SVM and cause false negatives on real images and videos. Using signed gradients also increases the time required for extracting features and training the SVM, because the feature vectors become longer (assuming the number of bins is 18) and therefore, the pipeline is slower. So, we think that one should stick to unsigned gradients for dashcam vehicle detection.

C. Linear SVM Parameters

The optimal Nu value was determined to be 0.09 with the trainAuto function using 10-fold cross-validation. However, although this value of the parameter minimized the testing error, we observed bad detection on real images and videos. After experimenting, we found out that for $Nu \approx 0.25$ the pipeline performs very good and the detections are very accurate. In this case the testing error is larger, but it improves the performance by significantly reducing the false positives on dashcam data. We assume that the vehicles and non-vehicles images from the dataset have more distinguishable HOG feature vectors, so smaller Nu value produces smaller error. On the other hand, sliding windows of a real image that contain or do not contain vehicles seem to have similar HOG feature vectors which requires larger misclassification cost or larger Nu value, thus reducing the number of false positives. So, the Nu value depends on the training data and the input images or videos.

D. detectMultiScale parameters

The optimal values of the detectMultiScale parameters depend on the input on which the detection is performed. The hitThreshold parameter is a threshold for the distance between the features and the SVM classifying plane. Setting its value too low will cause false positives and setting its value too high will lead to false negatives. In our experience, for best results the hitThreshold value should be between 1 and 2 depending on the finalThreshold value and the input image or video, when $Nu = 0.25$. The finalThreshold parameter is an NMS threshold. The function will classify a region as a vehicle if there are more positives in the region than this given threshold, otherwise it will classify the region as non-vehicle. Setting its value too low will cause false positives and setting its value too high will cause to false negatives. In our experience, for best results the finalThreshold value should be between 0.5 and 1.5 depending on the hitThreshold value and the input image or video, for $Nu = 0.25$. Figure 1 shows the results with hitThreshold set to 1.25 and finalThreshold set to 0.75, values that were found to give the desired results.

E. Real Data Performance

Although the described pipeline performs reasonably well on real dashcam images, its performance on real dashcam videos is not satisfying. It is unusual that there are lots of false positives in a video frame, while there are none in an image that seems identical to that video frame. This is the case because video frames contain artifacts which are caused by the application of lossy compression. HOG descriptors describe the shapes or the edges of an object, so the HOG feature vectors of those artifacts can be very similar to the ones of real vehicles. That will cause false positive misclassifications in unusual regions, like in the sky or the asphalt. As mentioned before, gamma correction, larger discrete derivative mask and Gaussian smoothing can be applied on videos to improve the detector's performance. However, vehicle detection in lossily compressed images and videos still suffers from false positives as they are not fully eliminated by tweaking these HOG parameters.

IV. DISCUSSION

The described approach performs reasonably well on images and achieves mediocre performance on videos. There are numerous proposed solutions for this problem. Unfortunately, tuning the HOG parameters is not one of them. Even if there are some optimal values for these parameters, performing detection on every video frame is computationally expensive. Bear in mind that those values will not be optimal for other videos. The right approach would be to implement vehicle

Figure 1. Vehicle detection in images with hitThreshold=1.25 and finalThreshold=0.75 with mean shift grouping

tracking, which is computationally inexpensive, rather than performing detection on every frame.

Some latest trends prefer using CNNs, RNNs or other deep learning approaches for these types of problems. We agree that these approaches are faster and more robust. However, our goal was to use classic image processing techniques instead of neural networks, because we wanted to get some perspective and experience in image processing and computer vision. So, although this approach cannot be used for real-time detection and is susceptible to false positives, it helps in understanding color spaces, image gradients, and SVM classifiers.

V. Conclusion and Future Work

We studied the influence of the color spaces on the performance and concluded that decorrelated and perceptual color spaces work best. We also studied the influence of the HOG parameters on the performance and concluded that in most cases, the optimal values of the HOG parameters for vehicle detection problems are the same as the proposed HOG parameters for human detection problems given in the original HOG paper [1]. HOG, SVM and detectMultiScale parameters are highly correlated and the choice of their values has profound effects on the performance of the detection. However, the performance of these parameters also depends on the input image or video and the goal should be finding the values for these parameters that generally work reasonably well with a lot of real images and videos and allow few misclassifications, instead of finding the perfect values for the parameters for only one image or video. The few misclassifications should be handled with other techniques.

We have shown that using HOG and linear SVM is a viable approach for vehicle detection in images, while it has some limitations for vehicle detection in videos. However, by using some simple techniques and extending the pipeline, this approach can easily overcome these limitations. Implementation of a vehicle tracking system is one of the future steps that will be considered for improving this pipeline. The false positives can be eliminated by checking whether the positive detections in a region are appearing in more consecutive frames. There are numerous HOG extensions and improvements that can be used, and SVMs with more complex kernels or modern and more sophisticated classification algorithms can be considered. These improvements can make the system resistant to artifacts and can provide overall better detection.

References

[1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection”, in *international Conference on computer vision & Pattern Recognition (CVPR’05)*, IEEE Computer Society, vol. 1, 2005, pp. 886–893.

[2] I. M. Creusen, R. G. Wijnhoven, E. Herbschleb, and P. de With, “Color exploitation in hog-based traffic sign detection”, in *2010 IEEE International Conference on Image Processing*, IEEE, 2010, pp. 2669–2672.

[3] L. Mao, M. Xie, Y. Huang, and Y. Zhang, “Preceding vehicle detection using histograms of oriented gradients”, in *2010 International Conference on Communications, Circuits and Systems (ICCCAS)*, IEEE, 2010, pp. 354–358.

[4] J. Arróspide, L. Salgado, and M. Camplani, “Image-based on-road vehicle detection using cost-effective histograms of oriented gradients”, *Journal of Visual Communication and Image Representation*, vol. 24, no. 7, pp. 1182–1190, 2013.

[5] R. Girshick, “Fast r-cnn”, in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[7] G. Rossum, “Python reference manual”, Amsterdam, The Netherlands, The Netherlands, Tech. Rep., 1995.

[8] G. Bradski, “The OpenCV Library”, *Dr. Dobb’s Journal of Software Tools*, 2000.

[9] I. P. G. at UPM, *Gti vehicle image database*, https://www.gti.ssr.upm.es/data/, 2011.

[10] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The kitti dataset”, *International Journal of Robotics Research (IJRR)*, 2013.

Journal of Emerging Computer Technologies
Araştırma Makalesi – Research Article
Received: *2021-01-04* | ***Reviewing:*** *2021-01-06* & *2021-02-10*| ***Accepted:*** *2021-02-15* | ***Online:*** *2021-02-16* | ***Issue Date:*** *2021-06-30*

Anlamsal Web ve Hastane Randevu Ontolojisi

Semantic Web and Hospital Appointment Ontology

Oğuz DÖNMEZ
Öğr. Gör. Dr.
Ege Üniversitesi
Tire Kutsan Meslek Yüksekokulu
Tire, İzmir, Türkiye
oguz.donmez@ege.edu.tr
ORCID: 0000-0002-9802-8743

Ahmet KAYA
(Corresponding Author)
Prof. Dr.
Ege Üniversitesi
Tire Kutsan Meslek Yüksekokulu
Tire, İzmir, Türkiye
ahmet.kaya@ege.edu.tr
ORCID: 0000-0002-6105-0787

***Özet*—** Ontolojiler, uygulamalar için çok önemli bir unsurdur. Yönetim bilişim sistemleri, e-ticaret uygulamaları gibi birçok alanda kullanılmaktadırlar. Ontolojiler ayrıca Web 3.0 olarak adlandırılan ve insanlara ek olarak bilgisayarlar ve sistemlerin okuyup anlayabildiği yeni teknoloji içinde de büyük rol sahibidir. Bu yeni Web için yeni ontoloji tabanlı uygulamalar ve bilgi mimarileri geliştirilmektedir. Bu yeni yapıyla birlikte yeni dillere ihtiyaç duyulmakta ve bu noktada ontolojilerin önemi ortaya çıkmaktadır. Bu çalışmada, web 1.0, web 2.0 ve anlamsal web olarak da geçen web 3.0 incelenip özetlenmiştir. Makineleri Web'e dâhil etmek için gerekli olan anlamsal Web'in yapıtaşlarından biri olan ontoloji konusunda bilgi verilmiştir. Ayrıca bu bilgiler ışığında geliştirilen bir ontoloji olan hastane randevu ontolojisi sunulmuştur.

***Anahtar Kelimeler*—** *Anlamsal web, www, web 3.0, ontoloji, hastane randevu ontolojisi*

***Abstract*—** Ontologies are a very important element for applications. They are used in many areas such as management information systems, e-commerce applications. Ontologies also play a major role in the new technology called Web 3.0, which computers and systems in addition to humans can read and understand. New ontology-based applications and information architectures are being developed for this new Web. In this study; web 1.0, web 2.0 and web 3.0, which is also called semantic web, are examined and summarized. Information is provided about ontology, one of the building blocks of the semantic web, which is necessary to incorporate machines into the web. Besides, hospital appointment ontology which is an ontology developed in the light of this information is presented.

***Keywords*—** *Semantic web, www, web 3.0, ontology, hospital appointment ontology*

I. GİRİŞ

Çeşitli teknolojik gelişmeler, özellikle de internetin gelişmesi ve genişlemesi ile birlikte veri kaynakları çeşitlendi. Veri kaynaklarının çeşitlenmesi yanında sayıları da büyük artış gösterdi. Bu artış aynı zamanda üretilen ve işlenen verinin boyutunda da büyük artışlara neden oldu. Bilgi, ülkeler ve toplum için teknolojik ve ekonomik gelişmenin temel öğesi haline gelmiştir. Bilgi, insanlığın var olmasından bugüne farklı tür ve düzeylerde var olmuş ve gelişimini sürdürmüştür. Bilgi özellikle sanayi devrimi ile birlikte ortaya çıkan hızlı kalkınma dalgası ile artış göstermiştir. Bilgi aktarıldığında önem ve anlam kazanacaktır. Böylece toplumun gelişmesine de katkı sunmaktadır. Bu nedenle bilginin saklanması ve kuşaklar arasında aktarılması büyük önem taşımaktadır. İnsanlığın başlangıcından bugüne taş gibi çeşitli materyaller üzerine kazınarak başlayan bilgi aktarımı çok çeşitli ortamlar ve şekiller ile aktarılmaya devam etmiştir. Son dönemde ise Bilgi ve İletişim Teknolojileri (BİT)'nin gelişmesi ve bu gelişimin bir sonucu olarak kullanıldığı alanların genişlemesi ve çeşitlenmesi ile bilgi ve bilginin kullanımı farklı bir boyut kazanmıştır. Bilginin kayıtlandığı ortamlar basılı ve elektronik olarak çeşitlendiğinden dolayı hızla büyüyen bilgi kitlesinin erişilmesi, denetlenmesi ve sınıflandırılması büyük bir meydan okuma olarak karşımza çıkmaktadır [1]

Günümüz dünyasında bilgiye ulaşmada kullanılan en temel araçlardan biri internettir. İnternet, milyonlarca bilgisayarın global olarak bağlandığı bir ağdır. Bu ağda herhangi bir bilgisayar diğer bilgisayarlarla iletişim kurabilir. World wide web ise internet üzerinde bilgiye ulaşmada bir yoldur. World wide web ile bilgiye ulaşıldıktan sonra bilgilerin linklerle bağlı olduğu ve metin, grafik, ses ve videoya sahip olabilen web sayfaları tarayıcıda yorumlanır ve bu şekilde kullanıcılar tarafından görüntülenebilir [2].

Eskiden web (web 1.0) bilgisel web olarak bilinirdi ve kullanıcı web sayfalarında sadece bilgi okuyup paylaşabilirdi. HTML belgelerinin yer aldığı donuk bir yapıya sahipti [3]. Daha sonraları web (web 2.0), kullanıcıların birbiriyle haberleşebildiği okuma ve yazma da yapılmaya başlanan ağ platformu haline dönüştü. Yani web 2.0 ile birlikte kullanıcıların aktif olarak katılabildiği etkileşim ve insan odaklı bir platform doğdu [3]. Günümüzde anlamsal web (web 3.0) ile birlikte web, insan yerine sistem tarafından okunan ve kategorize edilen bir dile dönüşmüştür. Tablo 1 [3], web'in evrimini göstermektedir.

Tablo 1. Web'in evrimi

Web 1.0	1995-2000 (Belge odaklı)
Web 2.0	2000-2010 (Etkileşimli ve insan odaklı)
Anlamsal Web (Web 3.0)	2010-2020 (Bilgi ve bilgisayar odaklı)

Dünya üzerinde üretilen bilginin tümünü tek bir ortamda toplama amacı ile ortaya çıkan anlamsal web, bu bilgi ve sürecin web üzerinde otomatik olarak bilgisayarlarca yönetilmesine imkan tanıyan bir uygulamadır. Anlamsal web ile bilgisayar insan, insan bilgisayar etkileşiminden bilgisayar-bilgisayar etkileşimine geçiş sağlanmıştır. Web 3.0 ile farklı cihazlar ile erişim imkânları artmış. Ayrıca arka planda çalışan yapay zekâ, makine öğrenmesi vb. yöntemler ile anlamsal olarak kullanıcı verileri işlenmekte cihazların karşılıklı olarak

veri alışverişi sağlanmaktadır. Bu durum, veri ve üretilen tüm bilginin web ortamında açıklamalar ile ilişkisinin kurulmasını gerektirmektedir. Buradaki kritik nokta elde edilen verilerin yönetilmesi amacıyla meta veri denilen bir kaynağın ya da verinin öğelerini tanımlayan bilginin oluşturulması ve bilgisayar sistemleri tarafından karmaşık sorgular ile ilgili bilgilere ulaşılmasıdır [4].

Anlamsal Web'in gerçeklenmesi için anahtar teknoloji ontolojilerdir. Bir uzmanlık veya uzmanlık sahasındaki kavramların, bu kavramlar arasındaki ilişkilerin ve hiyerarşinin tanımlandığı olguya ontoloji denir. Ontolojiler, uzun uğraşlar neticesinde son halini alır ve alanında uzman kişilerce incelenir ve tanımlanır. [5].

İkinci bölüm web 2.0'ın problemleri konusunda bilgiler içermektedir. Üçüncü bölümde web 2.0den sonra gelişmiş olan web 3.0 teknolojisi veya diğer adı ile anlamsal web konusunda bilgilere yer verilmiştir. Dördüncü bölümde ise anlamsal web için kritik öneme sahip ontoloji kavramı konusunda bilgiler verilmektedir. Son bölümde ise bu çalışma kapsamında geliştirilmiş hastane randevu ontolojisine yer verilmiştir.

II. WEB 2.0'IN PROBLEMLERİ

Ortaya koyduğu dinamik yapıya rağmen Web 2.0 bünyesinde birçok problemi barındırmaktadır [3]. Bu sorunlar aşağıdaki gibi listelenebilir:

a. Anahtar kelimeleri kullanarak sonuçları listeleyen Google ve Yahoo! gibi güçlü arama motorları Web 2.0 ile üretilmiş içeriklerden istenilen sonuçları vermekte zorluklar yaşamaktaydı.

b. Web 2.0 yapısı gereği bilgileri genelde düz metinler şeklinde sunduğundan dolayı bu bilgilerin bilgisayarlar ve diğer cihazlar tarafından anlaşılması zor idi. Bu bilgiler genelde sadece insanlar tarafından anlaşılabilmektedir.

c. Sunulan metin bazlı içeriklerin bilgisayarlarca okunması, işlenmesi, anlamlandırılıp diğer bilgisayarlar ile ilişkilendirilmesi veya başka ortamlara iletilmesi zordur.

d. Elde edilen bilgilerin arasında ilişki kurulmasının getirdiği zorluk nedeni ile bu bilgilerden yeni ve anlamlı bilgilerin otomatik olarak üretilmesi imkansız hale gelmektedir.

e. Güvenlik bakımından ise riskli bir ortam vadetmektedir. Bunun nedeni ise web 2.0 ile yayımlanan bilginin bilgisayarlar tarafından anlaşılması, işlenmesi ve ilişkilendirilmesi zor olduğundan dolayı içeriğin güvenlik açısından da takip edilmesini zorlaştırmaktadır.

Web 3.0 diğer adıyla anlamsal web, web 2.0'ın yukarıda zikredilen sorunlarını gidermek için ortaya çıktı.

III. ANLAMSAL WEB

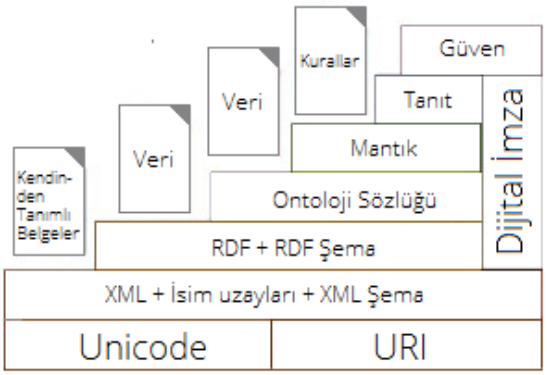
Tim Berners-Lee şu iki öngörüsü ile Web 'in geleceğe dönük konumunu ortaya koymuştur; "İlk olarak web, daha fazla birlikteliğin sağlanacağı bir ortam haline gelecektir, ikinci olarak da işlemlerin bilgisayarlar tarafından yürütülmesi ile web daha anlaşılır olacaktır" [6].

Anlamsal web ile insanların web 2.0'da olduğu gibi içeriği anlayıp kullanabildikleri bir ortam sunarken aynı zamanda bilginin açıkça tanımlanması, web sayfalarının yapısının bilgisayarlarca kolayca okunabilip işlenebilmesi gibi imkânlar sunmaktadır [7].

Yapay zekâyı temel alan bilgi işleme, sunma ve web teknolojileri ile bunun yürütülmesi anlamsal web'in ana öğelerinden birisi olmuştur. Başka bir deyişle anlamsal web, yapay zeka'nın web teknolojilerinin geliştirme ve uyumlaştırma sürecine eklenmesi ile gelişen bir alan olarak ifade edilebilir[8]. İyi şekilde tanımlanmış bir problemin çözümünde iyi tanımlanmış işlemleri ve yine mevcut tanımlanmış veriler üzerinde çözebilme yeteneğini makinenin anlayabileceği akıllı veri olarak ifade edebiliriz [8].

Anlamsal web ile birlikte bazı kavramlar ön plana çıkmaktadır. Bunlardan bazıları RDF, OML, OWL ve XML dilleridir. Bu teknoloji tümüyle birbiri ile bağlantılı basamaklara sahiptir. Bu isimlerini zikrettiğimiz dillerin XML dili dışındakilerde kendi sınıf modelleri vardır. Kaynak açıklama altyapısı (Resource Description Framework- RDF) olarak adlandırılan kavram anlamsal web 'in temel veri modelini oluşturur. RDF modeli, RDF/XML, N3, N-Triples ve Turtle gibi farklı değişim formatların temelini oluşturmaktadır. Ontolojiler, anlamsal web'in kalbini oluşturmaktadır. Web ontoloji dili (OWL), veriler ile tanımlamalar arasındaki bağlantıları kurmada kullanılmaktadır [8].

Anlamsal web, Şekil1'de katmanları görülebileceği gibi katmanlı bir yapıya sahiptir [9]. Anlamsal web katmanları, bir alt katman tarafından sunulan servisleri kullanırken bir üst katmana da yeni servisler sunar. Katmanlı bir mimari seçilmesi geliştirmenin daha esnek olabilmesini sağlamaktadır. Ayrıca tam ve eksiksiz bir sistem yerine eksiklikleri bilinen ve sonradan üzerine yeni eklemeler yapılabilen bir sistem olarak tasarlanmıştır. Bu nedenle, anlamsal web tasarımcısının hedefine ulaşana kadar evrimsel bir süreci izleyerek gelişecektir [10].



Şekil 1. Anlamsal Web katmanları [11; 12]

Anlamsal web katmanlarının en altındaki iki katman veri iletimini gerçekleştirmek için tasarlanmıştır. Unicode ve URI sayesinde kullanılan platform, uygulama ya da dil ne olursa olsun veri doğru bir şekilde taşınabilir. XML sayesinde de üst katmanlardaki uygulamalar arasında veri alış-verişi doğru bir şekilde yapılır [10].

RDF ve RDF Şema verilerin hakkında meta veri tanımlamaya yarar. RDF kullanarak web üzerindeki herhangi bir kaynak hakkında tanımlama yapılabilir. Bir web nesnesi hakkında RDF kullanılarak üçlülerden oluşan bir çizge

tanımlanarak onun hakkındaki meta veri gösterilir. RDF Şema ise RDF için kullanılmak üzere bir tip sistemi tanımlar [10].

Ontoloji katmanı, RDF ve RDF Şema ile tanımlanamayan daha karmaşık ilişkileri tanımlamak için yeni yapılar tanımlar. Ontoloji katmanının üzerindeki katmanlar ise çıkarsama için kural tanımlama, yapılan çıkarsamaların doğruluğunu ve güvenilirliğini denetlemek için kullanılır [10].

IV. ONTOLOJİ

Gruber, ontolojiyi bilgisayar bilimleri açısından en çok kabul gören şu tanımla ifade etmiştir [13]:

"Ontoloji, kavramsallaştırmanın açıkça belirtilmesidir."

Belli bir alanda soyut modeller oluşturmak kavramsallaştırma olarak ifade edilir. Bu soyut model, ilgili alandaki bilinen bütün bilgiyi hiçbir karanlık nokta kalmayacak şekilde tanımlamalıdır. Ontoloji, bilginin açıkça tanımlanmasını sağlayan ve tüm kavramları tanımlayan modeldir [10].

Ontoloji, felsefeciler tarafından kullanılan bir sözcüktür. Varlıkları ilişkileri ile birlikte tanımlamaktadır. Anlamsal web adlandırdığımız Web 3.0'ın temel bileşenidir. Web ontolojisi ile nesnelerin kurallı bir tanımı yapılmış ve bu şekilde ortak kelime ve anlamlar sunulmaktadır. Bu şekilde sunulan bilgi, web üzerinden bu bilgiye erişmek isteyenlere bilginin bir çerçeve ile sunulmasını sağlamaktadır [14].

Ontoloji şunları içerir [15]:

- Kavramlar
- Kavramların özellik ve nitelikleri
- Özellik ve nitelikler üzerine kısıtlamalar
- Bireyler

Ontoloji tarafından tanımlanan bilginin işlenmesi farklı bir araştırma konusu olarak karşımıza çıkmaktadır. Mantık, ontoloji ile tanımlanmış bilgiyi işlemede kullanılan araçtır. Ontoloji, mantıkla tanımlanan kurallardaki değişkenlerin yerine konan kavramları içerir. Ontoloji ve mantık birlikte eldeki bilgi üzerinde çıkarsama yapılması için bir sistem oluştururlar [10].

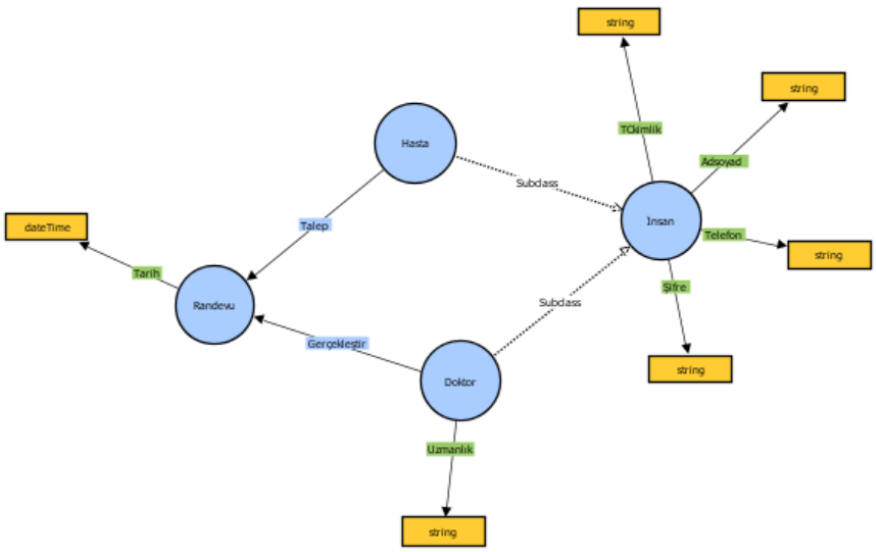
Araba kiralama, şarap yapımı, kara taşımacılığı, turizm, doğalgaz boru hattı bakımı gibi iş ve süreçler ontoloji geliştirmeye uygun uygulama alanları olarak sıralanabilir. Bu alanlar değerlendirildiğinde genel çerçevede sosyal bilimler, daha dar alanda ise işletme bilimleri alanına girmektedir [13].

Ontoloji ile birlikte bilginin kendi içinde ve diğer bilgiler ile ilişkilerini bilgisayar sistemlerinin anlayabileceği genel bir kavram çerçevesi çizilmektedir. Başka deyişle ontoloji geliştirme ve tanımlama işlemi bilginin tanımlanması standartlaştırmaktadır [6]. Ontoloji geliştirme güncel yazılım mühendisliği bilgisine sahip kişilerce yapılmalıdır ve gerektiğinde alanında uzman sosyal bilimci desteği de alınabilir.

Ontoloji bir dil olarak, modelin elektronik veya fiziksel ortamda saklanmasına bağlı olmaksızın kayıt altına alınmasını sağlayan bir araçtır. Ontoloji dilleri, temel amacı olan bilginin işlenmesi ve saklanması gibi konuları da dikkate alarak bir denge kurmalıdır. Çok karmaşık yapılar dahi bir ontoloji dili ile modellenebilir. Dilden gelen karmaşıklık, bu dil kullanılarak ortaya çıkarılmış ontolojilerin üzerinden çıkarsama yapılmasını zorlaştıracaktır [10]

OWL (Web Ontoloji Dili, Web Ontology Language) yaygın olarak kullanımda olan ve W2C tarafından 2002 yılında geliştirilen bir ontoloji dilidir. RDF (Kaynak Tanımlama Çerçevesi, Resource Description Framework) ise veri modelinin düzenlenmesini sağlamaktadır. Bu model bilgisayarlarca işlenecek verinin alamını temsil etmektedir. RDF modelinin genişletilmesi, kullanılacak nesneler, bu nesneler arası ilişki, özellik ve özelliklere atanabilecek değerler bakımından tanımlanması için RDFS (RDF Schema) gösterimi ortaya çıkarılmıştır [14].

RDF'ye ek olarak Avrupa Birliği tarafından OIL (Ontology Interface Layer) , ABD'nin desteği ile de DAML (DARPA Agent Markup Language) ontoloji dilleri tanımlanmıştır. Bunun yanında AB ve ABD tarafından geliştirilen DAML+OIL dili yaygın olarak kullanılmaktadır. Bu dil ilk olarak 200 yılında geliştirilmiş ve son sürümü 2001 yılında yayınlanmıştır [16].



Şekil 2. Hastane randevu ontolojisinin görsel gösterimi

V. GELİŞTİRİLEN HASTANE RANDEVU ONTOLOJİSİ

Bu çalışmada anlamsal web ve ontoloji kavramları incelenip hastane randevu ontolojisi tasarlanıp geliştirilmiştir. Geliştirme ortamı olarak Stanford üniversitesinin protégé ontoloji editörü kullanılmıştır [17]. Şekil 2'de bu ontolojinin bir gösterimi verilmektedir.

Şekil 2'de sınıflar, nesne özellikleri ve veri özellikleri gösterilmektedir. Doktor ve Hasta sınıfları İnsan sınıfının alt sınıflarıdır. İnsan sınıfı alanının veri özellikleri: TCkimlik, Adsoyad, Telefon ve Şifre'dir. Doktor sınıfı alanının veri özelliği Uzmanlık'tır. Randevu sınıfı alanının veri özelliği Tarih'tir. Randevu talep eden Talep ve randevuyu gerçekleştiren Gerçekleştir nesne özellikleri vardır.

VI. SONUÇ

Web 2.0'ın doğurduğu sorunları gidermek için anlamsal web'e ihtiyaç duyulmuştur. Bu çalışmada, anlamsal web ve ontoloji bilgileri ışığında bir hastane randevu ontolojisi tasarlanıp geliştirilmiştir.

Geliştirilen hastane randevu ontolojisinin sistemlerde kullanılmasıyla makinelerde insanlarla birlikte çalışarak, insanların yaptığı bazı işlemleri gerçekleştirebilir duruma gelecektir. Makinelerde, makinenin anlayabileceği akıllı verileri kullanarak problem çözebilir duruma gelecektir.

Önerilen ontoloji tasarımı, ekonomik, ölçeklenebilir ve sağlam bir Web tabanlı sağlık hizmetleri platformunun temelini oluşturabilir. Henüz ele alınması gereken kritik bir konu, kullanıcı gizliliğinin ve kişisel bilgilerinin korunmasıdır. Bu çalışma, geliştirilen hastane randevu ontolojisinin bir yazılım sistemine entegre edilmesiyle geliştirilebilir.

KAYNAKLAR

[1] Atılgan D. (2003)Bilimsel Bilgiye Erişimin Önemi Ve Türkiye'de Eğitim Araştırmaları Veri Tabanı. Akademik Bilişim, Çukurova Üniversitesi.

[2] Nath K., Dhar S. ve Basishtha S. (2014)Web 1.0 to Web 3.0-Evolution of the web and its various challanges. International Conference on Optimization, Reliabilty, and Information Technology (ICROIT), 86-90, Faridabad, NCR, India.

[3] Ege B. (2011) Yeni bilgi modelleme ve programlama felsefesiyle semantic web. Bilim ve Teknik dergisi, 2011.

[4] Emiroğlu B. G. (2009)Semantic web (anlamsal ağ) yapıları ve yansımaları. Akademik Bilişim, Harran Üniversitesi.

[5] Yıldırım A.ve Üsküdarlı S. (2015) , Bir Ontoloji ile Mikroblog Ortamlarının Modellenmesi ile, İçeriklerin Anlamsal Olarak Erişilebilir Hale Getirilmesi ve Sorgulanması. Akademik Bilişim, Anadolu Üniversitesi.

[6] Deconta M. C., Obrst J. L. ve SmithT. K.(2003) The semantic web, Wiley.

[7] Berners-Lee T., Hendler J. ve Lassila O. (2001) The Semantic Web. Scientific American, cilt 285, no. 5, 34-43.

[8] Gökçearslan Ş., (2011) Semantic Web (Web 3.0) ve Eğitim Amaçlı Kullanımı. Eğitim Teknolojleri Araştırmaları Dergisi, Cilt 2, Sayı 4.

[9] Berners-LeeT. (1999) Weaving the Web, Harper, New York.

[10] Öztürk Ö. (2004) Anlamsal Web İçin Bir Ontoloji Ortamı Tasarımı ve Gerçekleştirimi. Yüksek Lisans Tezi, Ege Üniversitesi.

[11] Seng, J. (2004) Digital Content Exchange https://james.seng.sg/2004/08/27/digital-content-exchange/

[12] Komesli, M. (2007) Cografi bilgi sistemlerinin OWL ve GML kullanarak anlamsal web ile bütünleştirilmesi. Ege Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi.

[13] Gruber T. R. (1993) A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5, 2, 199-220.

[14] Kurtel K. (2008) Web'in geleceği: anlamsal web. Ege Akademik Bakış, 8, 1, 205-213.

[15] Mishra, T. K. (2003). Semantic web. . Rapor, Indian Institute Of Technology Kharagpur.

[16] Akyokuş S. (2020) Anlamsal Web, Anlamsal Web Dilleri ve Araçları,http://www.edirnevdb.gov.tr/kultur/ppt/anlamsal_web_rdf_dc_owl.ppt, erişim tarihi: 24/06/2020.

[17] Protégé (2021) A free, open-source ontology editor and framework for building intelligent systems. https://protege.stanford.edu/

Journal of Emerging Computer Technologies
Research Article
Received: *2020-11-25* | ***Reviewing:*** *2020-11-26* & *2021-02-14* | ***Accepted:*** *2021-02-15* | ***Online:*** *2021-02-19* | ***Issue Date:*** *2021-06-30*

Edge Computing Security with an IoT device

Beyda Nur Kars
Dokuz Eylul University
Department of Computer Engineering
İzmir, TURKEY
beyda.kars@ceng.deu.edu.tr

***Abstract*— Information technologies are changing every aspect of human life day by day. In this context, Edge Computing, Internet of Things, Machine Learning and Big Data Analytics technologies are thought to be a part of this change. Edge computing aims to bring the computing power from the remote cloud environments to the endpoints/edges of networks. Thus, smart applications do not have to send all their data to the cloud and wait for the answers to come back over the same long route. Despite this advantage, there are security risks in the edge computing process. Encryption of information is of great importance especially for IoT devices to perform transactions safely. With a system established in this study, encrypted communication has been tried to be provided on IoT devices performing edge computing. In this way, it is aimed to make the communication secure. Arduino is used as an IoT device. In the encryption process, AES encryption is used with 128-bit and 256-bit key length.**

Keywords—Edge computing, security, AES, encryption, Arduino

I. INTRODUCTION

Edge Computing is an IT system designed to bring applications and computing capability as close as possible to the users or "objects" who need them. Edge Computing; mobile computing is driven by the reduced cost of computer components and the number of networked devices on the Internet of Things (IoT). Depending on the application, time-sensitive data in an Edge Computing can be processed by a smart device or sent to a medium server in a location close to the receiver. Edge computing is the optimization of the application's data with end-to-end encryption. That is, cloud computing turns it into a more sophisticated computing cloud architecture. Edge Computing is expected to play a serious role in the near future, especially in terms of information technologies. Edge Computing is referred to as end calculation in English. As it is known, the applications we use on smartphones have to deliver their existing data to the cloud, which is quite far away. Likewise, data is expected to return from this long journey. Edge Computing prevents this waste of time. Edge Computing, which can carry its computing power even to the extreme parts of the networks, brings serious relief in this sense. it cannot replace cloud computing or directly replace it. Edge Computing aims to process data from the ends to the cloud.

Despite all these positive aspects, Edge Computing can pose certain risks in terms of security. The fact that the architecture is distributed brings an increase in the number of vectors. Some vulnerabilities may also arise in the system. When more and more intelligence that an end customer has become more vulnerable, malware becomes infiltrations and vulnerabilities. On the other hand, licensing can force users to some extent, especially in terms of cost. Because each additional function is licensed separately, which naturally leads to higher costs.

Edge Computing brings advantages beyond classical architecture, as it optimizes resource usage significantly. First of all, the processors used in the related devices provide a relatively low power requirement and provide more efficient hardware security.

Mahadev Satyanarayanan, who is described as the father of Edge computing, explained how edge computing was born in his article [1]. According to Satyanarayanan, the limited resources of mobile devices have created a need for a platform that will perform the calculation process instead of these devices and have a more powerful resource. Cloud Computing, which promises much stronger resources (computing, memory and storage resources) compared to mobile devices, emerged to solve this problem. However, as latency-sensitive and bandwidth-hungry applications developed, gathering all computing power in the remote cloud environment started to cause problems. The main reason behind this problem is that the application packages have to pass through many routers controlled by many Internet Service Providers (ISPs) at different layers until they reach the cloud environment. Each router through which packets pass manifests itself as an increase in the Round Trip Time (RTT) of the packets of delay sensitive applications. In addition, the end-to-end delay values of the routes provided by routing protocols for packets can change dynamically due to ISPs and network problems and situations. In addition to all these, it is predicted that billions of IoT devices will connect to the internet and transfer data in the future. The fact that billions of devices send data to the cloud environment located at a single point in a remote location will inevitably lead to bandwidth bottlenecks. For all these reasons, it is clear that remote cloud environments cannot be a solution for latency sensitive and bandwidth hungry applications. Edge computing was born to solve this problem. Edge computing aims to prevent applications from being affected by long RTT times and bandwidth bottlenecks by moving cloud resources to the endpoints of networks.

II. RELATED WORKS

Confidentiality should be able to ensure that the data is only available to authorized users during the process and that it is not interfered by unauthorized people. Privacy is the most important security item in IoT because many devices can be integrated into IoT. It should be ensured that data received with a measuring device does not provide secure information to neighbouring devices. To ensure this privacy, advanced techniques and others, including key management mechanisms, should be developed and used [2].

It is essential to use wireless data transmission and to encrypt the information transferred between the nodes in order to keep it confidential. The most appropriate encryption algorithms and adequate key management systems are required to secure this data [3].

Wireless sensor networks have a large number of trust-based intrusion detection systems (IDS) that are used to defend against attacks. However, the effectiveness of IDS decreases in IoT due to the large amount of data produced in a short time. Meng et al. [4] proposed a Bayesian-based trust management method that incorporates traffic sampling into IDS under a hierarchical structure [5].

James King of Lulea University created an IoT network in his work in 2015 within the local network. Arduino used it as a gateway and collected and encrypted data from different devices and sensors. The encryption algorithm used is the Advanced Encryption Standard (AES) with both 128-bit and 256-bit key length [6]. Mahmudur Rahman, Bogdan Carbunar and Umut Topkara from Florida International University collects instant values from devices such as stopwatch, heart rhythm tracker and moisture meter used in fitness studies and transmits the IoT device to a web server remotely. In this study, FirstBeat is used as an encryption algorithm and Arduino Uno device is used as an IoT device. It also belongs to the IoT device from the webserver by taking the coordinate data, we can track the point we are at via GPS [6].

There is a need for special security mechanisms produced for low-resource devices such as IoT and wireless sensor network devices. It is necessary to use random number generators [7-11], authentication protocols [12-14] and some encryption algorithms that will be built-in in some of these devices.

III. PROPOSED WORK

In this section, we will briefly mention the target in the study. We will talk about the hardware materials that will be uses later and the flow of my work.

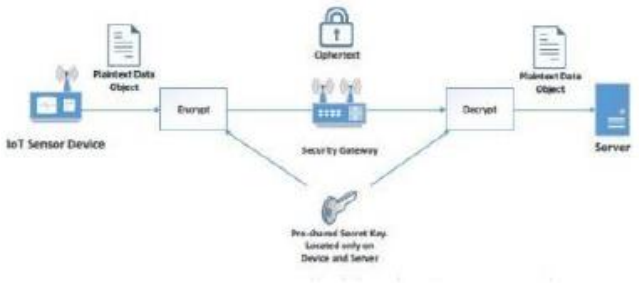


Figure 1. Encrypted data transfer

Our goal in this study is to encrypt data sent from IoT devices. This operation is shown in Figure 1. We used two Arduino devices in this study. The first one was used as a gateway device, the second one was used to get sensor data. DHT11 is used as the sensor to get data from the physical environment.

The data received from the sensor will be encrypted with AES 128 and sent to the network gateway. Thus, the data sent by IoT devices will be secured. Arduinos are also communicated among themselves. A wired connection has been made.

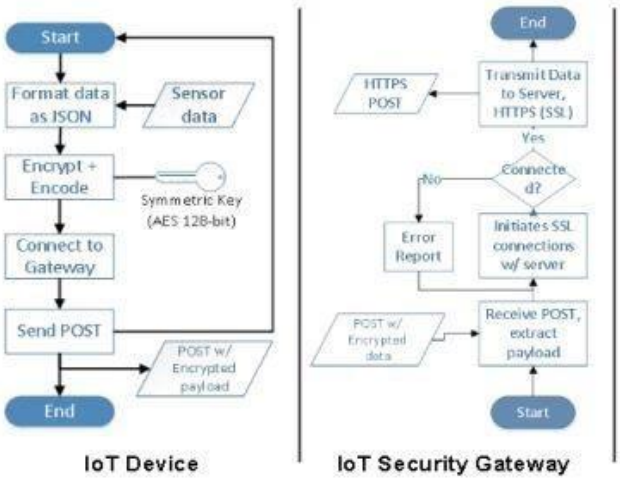


Figure 2. Workflow

In Figure 2, the workflow of operations of the two Arduino IoT device can be seen.

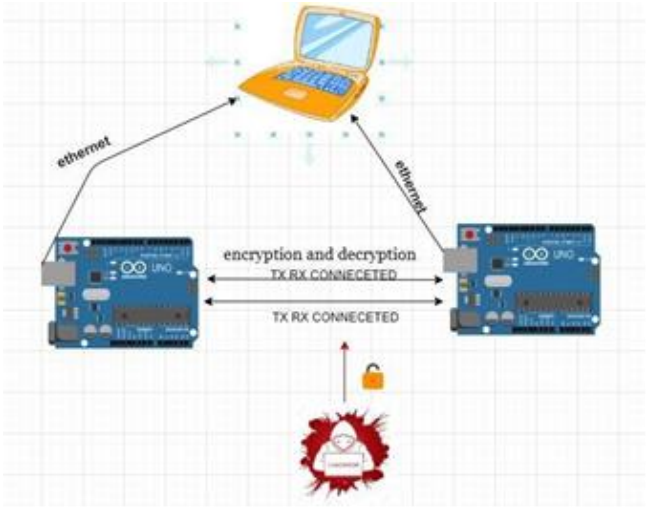


Figure 3. IOT Security Diagram

Figure 3 shows the relationship between all the necessary material in the workflow. Sensor data from Arduino 1 is sent to the laptop and encrypted there. And again from Arduino 2 to Arduino 1 (i.e. gateway is sent). Then Arduino 1 is decrypted.

The purpose of encrypting the data is to protect the data from malicious software while it is transmitted from one device to another.

IV. MATERIALS

As seen in the figures, we have two modules. The first one is the module with Arduino and sensor connection. The second one is Arduino, which will be used as gateway.

Figure 4 also has an Arduino module that allows us to receive data from the sensor.

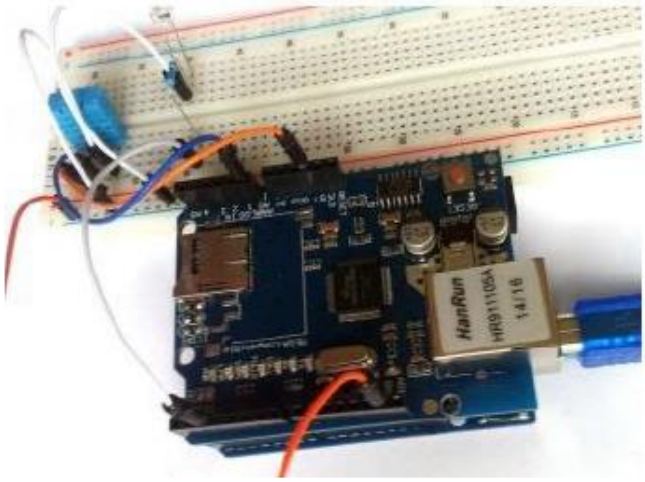


Figure 4. Arduino and Dht11 Connection

V. TEST RESULTS

Before talking about the test results in this part of the project, we will talk about the AES algorithm. We will give some information about the use of the AES algorithm in the working mechanism of this study.

The AES algorithm is a block cypher algorithm that encrypts 128-bit data blocks with 3 different key options, 128, 192 or 256-bit keys. The difference in key lengths makes the number of AES tour cycles differ. Since it is 128 bits, 10 rounds of processing are required. Since its length is 192 bits, 12 tours of processing is required. Because it is 256 bits, 14 rounds of operation are required [15].

Each round contains 4 different steps. These are as follows: byte replacement, row shift, column shuffling, and addition with a rotary switch. The data entered after 10 rounds is obtained as encrypted. The AES algorithm takes the key and passes it through certain processes, creating as many keys as the number of transactions. This number is 10 for 128-bit length. 10 different keys are created and the last key formed becomes the first key used to decrypt [16].

AES is a symmetric encryption algorithm, so the same key is used for both encryption and decryption. While deciphering the encrypted text, the operations performed while creating the encrypted text are applied in reverse.

The same key was used in encryption and decryption in the project as in Figure 5 and Figure 6.

```
uint8_t key[] = {0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15};
aes128_enc_single(key, data);
```

Figure 5. Code for the usage of the encryption key

```
uint8_t key[] = {0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15};
Serial.print("decrypted:");
aes128_dec_single(key, data);
```

Figure 6. Code for the usage of the decryption key

The Arduino's own library was used to encrypt the sensor data. The received data turns into encrypted text with the key we set. But since the sensor data that is received is a numerical value, the type conversion was made.

In the test part, we first load our code into our sensor module and continue to supply energy elsewhere by separating the module from the computer. Then we load the code into the module we will use as a gateway and provide the connection between the two Arduino devices. We ensure that encrypted data is sent to the gateway.

In Figure 7, circuits have all connections were shown.

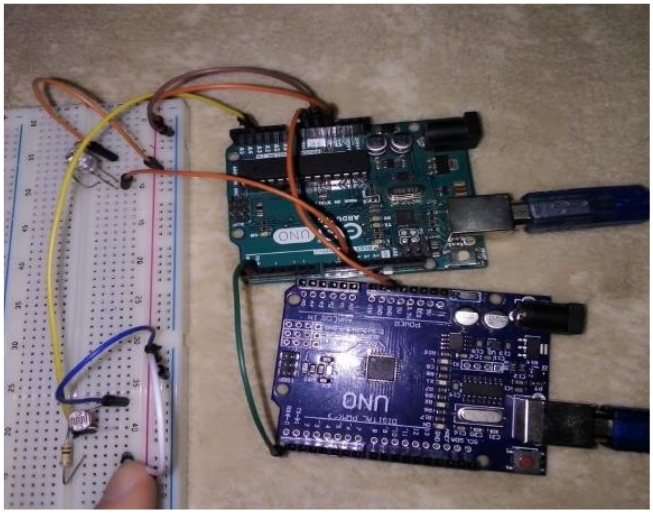


Figure 7. Hardware

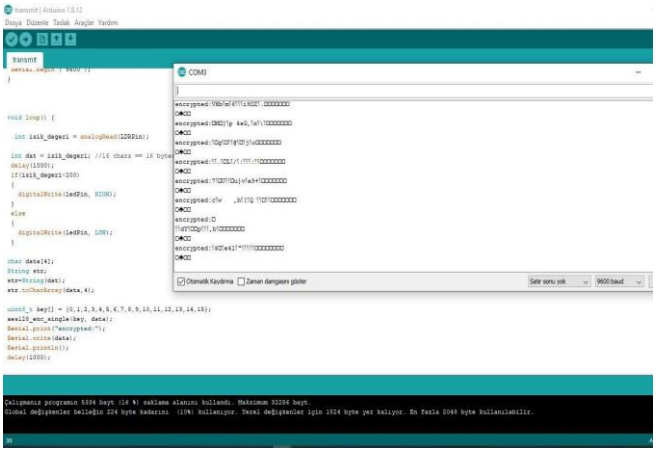


Figure 8. Transmit and Encrypted

Since there are two Arduino devices working as a receiver and a transmitter in the project, it was provided communication between the Arduino. This communication was made as a wired connection. Then the data received was encrypted and sent to the other party. The code part and serial port screen that is used for the receiver and the encryption process shown in Figure 8.

Figure 9 shows the encrypted version of the data received from the sensor. The code fragment in Figure 8 can be seen more clearly in Figure 10.

In Figure 11, there are pieces of code for both transmit and receive, that is decryption. Using these codes and operations, data communication is carried out by encrypting with AES. In this way, communication security will be provided for the edge computing process.

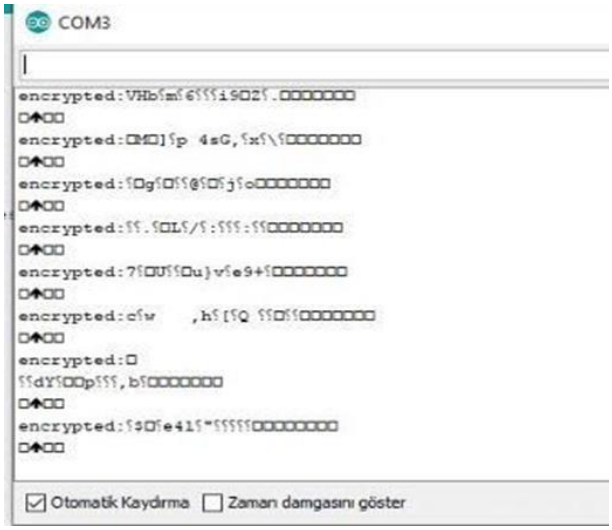


Figure 9. Receive and Decrypted

```
#include <AESLib.h>
int ledPin = 10;
int LDRPin = A3;
void setup() {
 pinMode(ledPin, OUTPUT);
 Serial.begin ( 9600 );
}
void loop() {
 int isik_degeri = analogRead(LDRPin);
 int dat = isik_degeri; //16 chars == 16 bytes
 delay(1000);
char data[4];
String str;
str=String(dat);
str.toCharArray(data,4);
uint8_t key[] = {0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15};
aes128_enc_single(key, data);
Serial.print("encrypted:");
Serial.print(data);
Serial.println();
delay(1000);
}
```

Figure 10. Arduino code for encryption

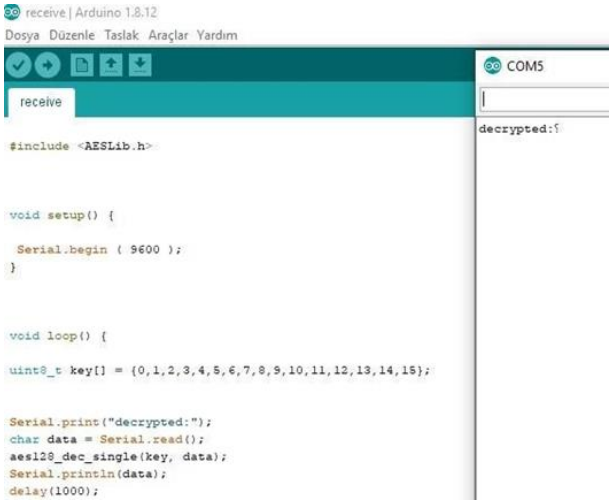


Figure 11. Receive and Decrypted

VI. Conclusion

In this study, an edge computing system was established with two Arduino devices. One of the devices was used as a gateway. The other is used to collect the data obtained from the sensor. The collected data is encrypted with AES encryption algorithm to ensure secure communication of data. Encryption methods with 128-bit and 256-bit key lengths have been tried for the encryption process. The AES library on Arduino was used for encryption. With this study, the data was sent encrypted and the data was reconstructed by decrypting the other side. Different encryption methods and different security mechanisms may be tried in the future.

References

[1] Satyanarayanan, M. (2017). Edge Computing. IEEE Computer, 50(10), 36-38.

[2] Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., & Zhao, W. (2017). A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. IEEE Internet of Things Journal, 4(5), 1125-1142.

[3] Bull, P., Austin, R., Popov, E., Sharma, M., & Watson, R. (2016, August). Flow based security for IoT devices using an SDN gateway. In 2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud) (pp. 157-163). IEEE.

[4] Meng, W., Li, W., Su, C., Zhou, J., & Lu, R. (2017). Enhancing trust management for wireless intrusion detection via traffic sampling in the era of big data. Ieee Access, 6, 7234-7243.

[5] Wang, T., Zhang, G., Liu, A., Bhuiyan, M. Z. A., & Jin, Q. (2018). A secure IoT service architecture with an efficient balance dynamics based on cloud and edge computing. IEEE Internet of Things Journal, 6(3), 4831-4843.

[6] Akkuş, S. (2016). Gömülü sistem tabanlı, kriptolu TCP/IP veri haberleşmesi uygulaması.

[7] Cabuk, U. C., Aydın, Ö., & Dalkılıç, G. (2017). A random number generator for lightweight authentication protocols: xorshiftR+. Turkish Journal of Electrical Engineering & Computer Sciences, 25(6), 4818-4828.

[8] Kösemen, C., Dalkılıç, G., & Aydın, Ö. (2018). Genetic programming-based pseudorandom number generator for wireless identification and sensing platform. Turkish Journal of Electrical Engineering & Computer Sciences, 26(5), 2500-2511.

[9] O'Neill ME. PCG: A family of simple fast space-efficient statistically good algorithms for random number generation.Available online at http://www.pcg-random.org/pdf/toms-oneill-pcg-family-v1.02.pdf

[10] Aydın, Ö., & Dalkılıç, G. (2018, July). A hybrid random number generator for lightweight cryptosystems: xorshiftLplus. The 3rd International Conference on Engineering Technology and Applied Sciences (ICETAS).

[11] Aydın, Ö., & Kösemen, C. (2020). XorshiftUL+: A novel hybrid random number generator for internet of things and wireless sensor network applications. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 26(5), 953-958.

[12] Armknecht F, Hamann M, Mikhalev V. Lightweight authentication protocols on ultra-constrained RFIDs-mythsand facts. In: International Workshop on Radio Frequency Identification: Security and Privacy Issues; 2015. pp.1-18. Springer, Cham.

[13] Lee JY, Lin WC, Huang YH. A lightweight authentication protocol for internet of things. In: 2014 InternationalSymposium on Next-Generation Electronics (ISNE); New York, USA; 2014. pp. 1-2.

[14] Aydın, Ö., Dalkılıç, G., & Kösemen, C. (2020). A novel grouping proof authentication protocol for lightweight devices: GPAPXR+. Turkish Journal of Electrical Engineering & Computer Sciences, 28(5), 3036-3051.

[15] Şeker, Ş.E. (2009). AES ve Rijndeal Şifreleme. *Bilgisayar Kavramları*. Retrieved from http://bilgisayarkavramlari.com/2009/06/03/aes-ve-rijndeal-sifreleme/ , Accessed date : 25.11.2020

[16] Surian, D. (2006). Algoritma Kriptografi AES Rijndael. TESLA Jurnal Teknik Elektro UNTAR, 8(2), 97-101..

Journal of Emerging Computer Technologies
Research Article
Received: *2020-12-09* | ***Reviewing:*** *2020-12-11* & *2021-01-18*| ***Accepted:*** *2021-01-26* | ***Online:*** *2021-02-20* | ***Issue Date:*** *2021-06-30*

Secure Connection between Google Home and IoT Device

Ekrem YİĞİT
Dokuz Eylul University
Department of Computer Engineering
İzmir, TURKEY
ekrem.yigit@ceng.deu.edu.tr

***Abstract*— This article presents a more secure connection between the NodemCU and Blynk, using the AES algorithm. it is aimed to prevent a vulnerability in the connection of Google Home devices with IoT during the Blynk IoT connection phase.**

Although the Blynk application offers a personal key, the network information linked in the memory of the physical device is vulnerable. The data is placed in the software by encrypting beforehand, and the connection is provided by decrypt during the connection. This study contributes to the prevention of security weaknesses caused by keeping software data on NodemCU as Plain Text in physical memory

Keywords—AES, IFTTT, IoT, Connection, Secure, Memory, EEPROM Introduction, Home Mini

I. INTRODUCTION

Google Home is a device developed by Google that can be customized both with its own functions and with many devices connected via Wi-Fi. It offers an interface that is managed by voice commands and can be customized.

Approximately 6 weeks after the release of Google Home devices, the logs kept on the devices were checked and it was determined that the device recorded the sounds as a log hacker can use these data to penetration attempts or fishing attacks. Researchers have found that smart speakers can be hacked with the help of laser-powered "light commands." Researchers suggest smart speaker makers can fix this vulnerability by adding a light shield around the microphone or using two different microphones on opposite sides to listen to voice commands.

The project is to operate the IoT device with safe commands sent from google home. In the project, the main target is to connect to the cloud from the google home device via IFTTT, then to get the light effect by taking data from the cloud using the NodemCU IoT device and Blynk.

NodemCU is a circuit board with high quality ESP8266 Wi-Fi module that can be programmed from Arduino IDE application and can communicate easily. It is aimed to provide cloud interaction by communicating with this device over the internet by making web transactions thanks to HTTP libraries. IFTTT (if this then that), If this happens, do it. It is an abbreviation of one of the simplest code phrases. This assign device to be triggered and the that assign action that works on device that is triggered. Webhooks service is a IFTTT service that connect to Blynk and IFTTT on Cloud Service. Webhooks is used to send requests to the URL we enter. Blynk server information and the TOKEN information used are entered as URLs and connection is provided and control is provided via IFTTT. In this URL, the pins on the device are entered and the controls of various pins are provided. Finally, my project is to identify vulnerabilities in communication on Blynk and IFTTT connections and to use these two components effectively in the safest possible way.

II. RELATED WORKS

The interaction of end-users with online IoT devices has been increasing for the last 10 years and with the increasing interactions, the usage of these IoT devices has started to be observed beyond the intended use. IoT devices need lightweight and new security mechanisms like random number generators [1-3], authentication protocols [4,5] and privacy protected connection methods. IFTTT enables users to use their home assistant devices heterogeneously on multiple platforms with IoT via the internet connection tool. With the widespread use of this environment, users are satisfied with the increase in interaction and they started to load more work on their systems.

In order to meet the increasing demand for products, studies on the field of security in a competitive environment have been quite incomplete. According to the researches carried out today, more than 50% of the 19323-connection established over home assistant devices did not find any security factor.

While the picture taken from the phone is in a private structure, the smart home device translates this picture information into a Public form. In the scenario described, the hacker sees the system vulnerabilities related to IFTTT, which enables the connection of the smart home device to the phone, and the Home Assistant will also make private content public and make confidential data accessible [6].

Security problems related to 45 smart home appliances and IoT devices connecting with them are discussed. It is determined that the most common vulnerability is users. Users endanger their own security by using Home Assistants with different levels of security for the same or similar jobs or IoT devices attached to them. In this research [7], which presents this vulnerability as the simplest and most effective solution, the use of multiple IoT and Home Assistant was shaped and the relational relationships were explained to the users.

It is not very difficult to steal users' IFTTT accounts with various online attack methods. Attackers who will change the authorizations made through this IFTTT account on another device for their own malicious purposes can cause damage as much as cyber-attacks with increasing device interactions [6].

This framework [8], developed to prevent a bad interaction via IFTTT, inserts the MAC address of the device to which the user connects to the IoT device to a PUF at the beginning,

and returns the output it receives as OTP, allowing access only from the devices authorized by the user at the beginning.

The most important prioritization on the market is always the cost of IoT devices and the platforms actually used. In costing-priority product planning, security is placed in the background. Extra cost or does not require any hardware. The answer that best meets these needs is the AES algorithm, which is also described in another research [9]. Offering both fast and secure service at once, AES also does not cause negative effects on IoT device performance [10].

Due to the lack of encryption in the transmission of data from IFTTT to IoT device of systems that are tried to be created with IFTTT account security and device connection methods seen in various studies, there are also vulnerabilities in leaking many data by logging them, so AES encrypt as an effective and cost effective method. With the method, both the data coming from IFTTT and Google Home Assistant is encrypted, and then the data is decoded without coming to the IoT device and secure transmission of the data [9].

Çepik et. al presented a study in 2020 about security vulnerabilities on the connection between Google Home and IoT device. They tested if this communication is vulnerable to an NTP attack [11].

III. PROPOSED WORK

The 2 basic devices of the project are the Google Home Mini and NumedCU IoT device. The microphone section on the device must be active. Make AES decryption on NumedCU, AES library has been added to Arduino IDE. The necessary encryption functions can be realized with this library. Password and SSID are necessity to connect NumedCU via Blynk. SSID and password If the information is added directly into the post to provide a link, a security flaw occurs here. While connecting, this two information in the software pose security risks. The attacker can discover certain data on NodemCU by reverse engineering methods via EPPROM.

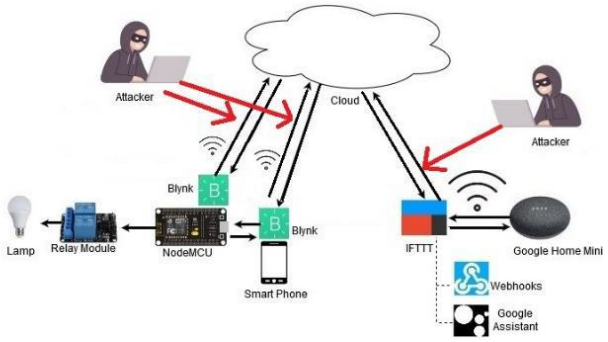


Figure 1. Representation of Possible System Vulnerabilities

All data entered while communicating with Blynk in NodemCU is kept in EEPROM. It is seen in the research that it is possible to access the data in EEPROM physically via a USB [12]. The attacker, who accesses the data via USB, can easily access the internet information on which the device is used, since no hardware measures are taken.

In another study to prevent EEPROM access in terms of hardware, by adding extra parts, the hardware vulnerability of the attackers was prevented but the software vulnerabilities could not be blocked [13].

The attackers' access to the information of the internet connection is a danger to all devices on that connection. To eliminate these risks, SSID and password will be added by encrypting with 128bit AES key in this study and this encrypted data will be provided by decryption while the connection is provided. 128Bit 16ASCII characters will be used for the key.

Reverse engineering of Arduino memory has been handled in the study in which it is possible to access information in the software [14]. The key size of the AES algorithm has a direct effect on performance in various algorithm encounters. When high security is targeted, performance/security threshold has been seen in the examinations; AES is more successful in this rate [15]. After a secure connection is established, the command to turn on the bulb coming from the cloud will be realized.

IV. TEST RESULTS

The main goal of the project is to add the user-known internet information to the program in a previously encrypted form and to provide an encrypted internet connection. Communication with four devices and a common internet connection has been tested.

The commands set via IFTTT to the Google home device were given as HTTP request, and the connection with the IoT device was tested and the lamp was successfully controlled. Button on / off communication to Blynk port 0 was transmitted in HTTP request format.

```
EkremEv
Encrypted: 2nqyFb+a9uGmJw52SAZtSg==
Decrypted: EkremEv
EkremEv
Encrypted: feRuuTz/lmkPbVaUh0i3Fg==
Decrypted: EkremEv
EkremEv
Encrypted: I6VVDN72z+Erlbtv2jvAyA==
Decrypted: EkremEv
EkremEv
Encrypted: A0NCgvzNjquSq9YR+NddSA==
Decrypted: EkremEv
EkremEv
Encrypted: C+AEieiqT2bgLmuZc6ECvg==
Decrypted: EkremEv
EkremEv
Encrypted: k8x53C+MKWj5Fk2PQpVUQA==
Decrypted: EkremEv
EkremEv
Encrypted: Nku4H7Q6t4aSgNfw79ODkA==
Decrypted: EkremEv
EkremEv
Encrypted: 6HZEKJk4D/zL3uUUfTiBeg==
Decrypted: EkremEv
EkremEv
Encrypted: 7YtI/cP7RrHtxr3SZv26Dw==
Decrypted: EkremEv
EkremEv
Encrypted: 2oPq27mZI6tX1NS+NBMakQ==
```

Figure 2 Arduino IDE Serial Port Output

On the security side of the application, Blynk internet connection has been tested to properly encrypt and decrypted data. As seen in the figure below, the software used on the IoT device processes the data properly. When performing the tests,

the IV has been re-created while performing each encryption in 16 bytes.

The connection made with the SSID and password of the user, which was previously encrypted, was successfully implemented. These devices, which generally communicate with the cloud, have become a necessity in order to provide sufficient trust. As mentioned in many studies like [16], NodemCU does not have sufficient security, so it is coded into additional security algorithms.

V. CONCLUSION

As a result, data in the NodemCU memory is not available in a meaningful way. All of important data are kept in encrypted form in memory and communication is provided in encrypted form. Even if the attacker has accessed to NodemCU via USB or software, the attacker will see data encrypted with AES, which he cannot make sense of. This prevents the device from being physically hijacked or the attacker from exploiting vulnerabilities.

REFERENCES

[1] Cabuk, U. C., Aydın, Ö., & Dalkılıç, G. (2017). A random number generator for lightweight authentication protocols: xorshiftR+. Turkish Journal of Electrical Engineering & Computer Sciences, 25(6), 4818-4828.

[2] Aydın, Ö., & Dalkılıç, G. (2018, July). A hybrid random number generator for lightweight cryptosystems: xorshiftLplus. The 3rd International Conference on Engineering Technology and Applied Sciences (ICETAS).

[3] Aydın, Ö., & Kösemen, C. (2020). XorshiftUL+: A novel hybrid random number generator for internet of things and wireless sensor network applications. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 26(5), 953-958.

[4] Lee JY, Lin WC, Huang YH. A lightweight authentication protocol for internet of things. In: 2014 InternationalSymposium on Next-Generation Electronics (ISNE); New York, USA; 2014. pp. 1-2.

[5] Aydın, Ö., Dalkılıç, G., & Kösemen, C. (2020). A novel grouping proof authentication protocol for lightweight devices: GPAPXR+. Turkish Journal of Electrical Engineering & Computer Sciences, 28(5), 3036-3051.

[6] Milijana Surbatovich, Jassim Aljuraidan,Lujo Bauer,Some Recipes Can Do More Than Spoil Your Appetite: Analyzing the Security and Privacy Risks of IFTTT Recipes" WWW '17: Proceedings of the 26th International Conference on World Wide WebApril 2017

[7] OmarAlrawi,Chaz Lever,ManosAntonakakis,FabianMonrose,Security Evaluation of Home-Based IoT Deployments, 2019 IEEE Symposium on Security and Privacy, 19-23 May 2019

[8] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[9] Bennet Praba1,Shivam Kumar,Ankur Saxena, Sourabh Patel "Securing Assistant Based Home Automation using AES Algorithm" Department of Computer Science and Engineering SRM Institute of Science and Technology, Chennai, India,2019

[10] Ross Mcpherson James Irvine, "Using Smartphones to Enable Low-Cost Secure Consumer IoT Devices," Department of Electronic and Electrical Engineering, Royal College Building, Glasgow G1 1XW, U.K..

[11] Çepik, H., Aydın, Ö., Dalkılıç, G. (2020) Security Vulnerability Assessment of Google Home Connection with an Internet of Things Device. In:7th International Management Information Systems Conference, İzmir, Turkey (9-11 December 2020.

[12] Kingpin, "Attacks on and Countermeasures for USB Hardware Token Devices," Proceedings of the Fifth Nordic Workshop on Secure IT Systems, 2000

[13] Grand, Joe. "Practical secure hardware design for embedded systems." Proceedings of the 2004 Embedded Systems Conference, San Francisco, California. 2004.

[14] Torroja, Yago, et al. "A serial port based debugging tool to improve learning with arduino." 2015 Conference on Design of Circuits and Integrated Systems (DCIS). IEEE, 2015.

[15] Kumar, M. Anand, and S. Karthikeyan. "Investigating the efficiency of Blowfish and Rejindael (AES) algorithms." International Journal of Computer Network and Information Security 4.2 (2012): 22.

[16] Utpala, Kuchi NSSSS, N. Suresh Kumar, K. Praneetha, D. Hema Sruthi, and K. Sai Avinash Varma. "Authenticated IoT Based Online Smart Parking System with Cloud." Pramana Research Journal 9, no. 4 (2019).