VSRD-TNTR

VSRD INTERNATIONAL JOURNAL OF TECHNICAL & NON-TECHNICAL RESEARCH

e-ISSN: 0976-7967, p-ISSN: 2319-2216

SPECIAL ISSUE

VOLUME XIV MAY 2023

EDITORS: Dr. H. Lilly Beaulah, Dr. R. Reka, Prof. C. Vinoth
Prof. L. Vinitha Sree, Prof. M. Sathya

7th National Conference on

"Artificial intelligence and Internet of Things on Management, Science and Technology"

AAIOT 2023

Organised By

MAHENDRA COLLEGE OF ENGINEERING

Salem- Chennai NH 79, Minnampalli, Salem, Tamilnadu, India

Web: www.mahendra.org, Ph. 0427-2482884/65423333

● Volume XIV ● Issue (Special Issue) ● May 2023

VSRD INTERNATIONAL JOURNAL OF
TECHNICAL AND NON-TECHNICAL RESEARCH
e-ISSN: 0976-7967, p-ISSN: 2319-2216

NATIONAL CONFERENCE ON "ARTIFICIAL INTELLIGENCE AND INTERNET OF THINGS ON MANAGEMENT, SCIENCE AND TECHNOLOGY" (AAIOT 2023)

EDITORS

Dr. H. Lilly Beaulah
Dr. R. Reka
Prof. C. Vinoth
Prof. L. Vinitha Sree
Prof. M. Sathya

ORGANISED BY

MAHENDRA COLLEGE OF ENGINEERING

Salem- Chennai NH 79, Minnampalli, Salem, Tamilnadu, India.
Web: www.mahendra.org
Ph. 0427-2482884/65423333

● **Volume XIV ● Issue (Special Issue) ● May2023**

VSRD INTERNATIONAL JOURNAL OF TECHNICAL AND NON-TECHNICAL RESEARCH
e-ISSN: 0976-7967, p-ISSN: 2319-2216

NATIONAL CONFERENCE ON "ARTIFICIAL INTELLIGENCE AND INTERNET OF THINGS ON MANAGEMENT, SCIENCE AND TECHNOLOGY" (AAIOT 2023)

Printed &Published by:
VSRD International Journals
A Research Division of Visual Soft (India) Private Limited

Printed &Bound in India

VSRD INTERNATIONAL JOURNALS
A Research Division of Visual Soft (India) Pvt. Ltd.

REGISTERED OFFICE
154, Tezabmill Campus, Anwarganj, KANPUR – 208 003 (UP) (INDIA)
Mob.: +91 99561 27040 || Web.: www.vsrdjournals.com || Email: vsrdjournal@gmail.com

MARKETING OFFICE
340, First Floor, Adarsh Nagar, Oshiwara, Andheri(W), MUMBAI – 400 053 (MH) (INDIA)
Mob.: +91 99561 27040 || Web.: www.vsrdjournals.com || Email: vsrdjournal@gmail.com

PREFACE

ABOUT THE CONFERENCE

7TH NATIONAL CONFERENCE ON APPLICATION OF ARTIFICIAL INTELLIGENCE AND INTERNET OF THINGS ON MANAGEMENT, SCIENCE AND TECHNOLOGY

The conference is "**Systematic and Comprehensive Investigation and Exploration of Nature's Causes and Effects**" in Application of Artificial Intelligence and Internet of Things based Science and Technology that reflects multidisciplinary nature that will provide a venue where convergence and innovation that will be shared through interdisciplinary integrations. This Conference aims to bring together leading Academicians, Scientists, Researchers and Scholars to exchange and share their experiences, research results on all aspects of Internet of Things and AI. Researchers will present and discuss the most recent innovations, trends, and concerns as well as practical challenges encountered in Artificial Intelligence for IoT. This conference is to encourage and assist the professionals engaged in the above fields to maintain the integrity and competence of the profession foster a sense of partnership amongst the International Professionals.

ABOUT THE INSTITUTION

With the noble intent of Educating the under privileged in the Salem and Namakkal Districts, Mahendra Educational Trust (MET) was founded in the year 1978, by the great Educationist and Philanthropist Shri. M.G. Bharath Kumar. Mahendra College of Engineering, established in the year 2005. The college is affiliated to Anna University, Chennai and duly recognized by the AICTE, New Delhi. It offers Undergraduate Programmes in Engineering and Technology and Five Post Graduate Programmes in the field of Engineering and Technology, College has been accredited with NAAC and 2(f) & 12(B) status from UGC, New Delhi. This college seeks to impart value based Technical Education to its students in order to meet the growing needs of such Technocrats and Entrepreneurs required in industries and business establishments. The college is situated in Salem. Mahendra College of Engineering presently offers UG /PG Degree Courses

ABOUT THE DEPARTMENT

The Department of CSE, established in the year 2005. It has grown excellently in academic and research work. The departments offer UG and PG courses. The departments are well equipped with the state of the art computing and internet facilities supported by experienced and well qualified faculty members who witnessed the tremendous growth in academics and research. Major research areas include Image Processing, Multimedia, Data Mining, Cloud Computing, Network Security and Wireless Networks. It has vibrant Associations like CSI, ISTE, and IEEE. The department aims at developing intellectually alert scientifically progressive globally competent and dynamic young IT professionals

ADVISORY COMMITTEE

Er. ARAVIND RAMAKRISHNAN, Project Analyst-Ford Motor Company, Auburn Hills, Michigan, USA
Er.PAGALAVAN KRISHNAMOORTHY, Manager-SW development at Cisco Systems, San Jose, California, USA
Er.BALU NARAYANAN, IT Manager AT&T, Naperville, Illinois, United States
Dr. P. S. SRIRAJ, Director of Urban Transportation Centre, University of Illinois at Chicago, USA.
DEVNATH VAIJA RAJARAM, Product Development Engineer, Ford Motor Company, Dearborn, MI-USA.
SHARATH KUMAR GUNASEKARAN, Data Analyst, Brisbane, Queensland, Australia
Er. GAYATHRI JEYAPAL, Senior Developer/SDk, Technical Writer, New Zealand
Dr. MURUGESAN RANGANATHAN, Chairman, Ranganathan Educational Institutions. Coimbatore, Tamil Nadu, India.
Dr. A.K. NATESAN, Founder & Chairman, Excel Group Institutions, Namakkal, Tamilnadu, India
ER. ASHWIN SEKAR, Software Design Engineer, Bengaluru, Karnataka, India
Er.VISHNU S PRASAD, Director, Saltu Creative Suite Private Limited, Alappuzha, Kerala, India
Mr. S. SATHISKUMAR, Founder, Director, CEO., Imagecon Academy, Salem, India
Er. RAM KUMAR ARUMUGAM, Consultant at Delomitte, Bangaluru, Karnataka, India
RUPESH KUMAR SINGH, Senior Project Manager – Infosys, Odisha, India

AAIOT2023

OR G AN I Z ING C OMMI T T E E

HONOURABLE CHAIR

Thirumigu. M.G. BHARATHKUMAR
Founder & Chairman, Mahendra Educational Trust, Namakkal-Salem Dt

Srimathi. B. VALLIAMMAL
Secretary, Mahendra Educational Trust, Namakkal-Salem Dt

C H I E F -P A T R O N

Dr. R. SAMSON RAVINDRAN
Executive Director, Mahendra Group of Colleges, Namakkal-Salem Dt

P A T R O N

Dr. N. MOHANA SUNDARA RAJU
Principa, Mahendra College of Engineering, Minnampalli, Salem

C O N F E R E N C E- C H A I R

Dr. A. SUPHALAKSHMI
Professor & head / CSE, Takshashila University, Ongur, Tindivanam

C O N V E N E R

Dr. H. LILLY BEAULAH
Prof., & Head / CSE, Mahendra College of Engineering, Minnampalli, Salem

COORDINATORS

Dr. R. Reka, Asso.Prof. /CSE
Ms. L. Vinitha Sree, AP/CSE, M. Sathya, AP/CSE
Mahendra College of Engineering, Minnampalli, Salem

ORGANISING COMMITTEES

Mr. M. Jenolin Rex, AP/CSE, Mrs. V. Deepa, AP/CSE, Mrs. V. Nisha Devi, AP / CSE
Mrs. A. Indhuja, AP /CSE, Mr. M. Anandraj, AP /CSE, Ms. M. Gayathri, AP /CSE
Mr. A. Amjath, AP /CSE, Ms. V. Dhanakodi, AP /CSE,
Mr. C. Vinoth, AP /CSE, Mr. Riyaz, AP/CSE
Mahendra College of Engineering, Minnampalli, Salem

STUDENT COORDINATORS

Mr. U.S. Parvez Musharaf, Final /CSE, Ms.T. Prithisha, Third/CSE, Ms. K. Srinithi, Second /CSE
Mahendra College of Engineering, Minnampalli, Salem

CONTENTS

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

New Policy-Based Xss Protection Mechanism for Browsers: JSCSP

SUKUMAR C[1], TAMIZHARASAN M[2], BAVISHA G[3], SANMUGA PRIYA S[4]
[1]Assistant Professor, [2,3,4]Student
[1,2,3,4]Department of CSE, Annapoorana Engineering College, Salem, Tamil Nadu, India

ABSTRACT

The W3C organization advises web service providers to use the computer security standard known as Content Security Policy to reduce cross-site scripting assaults (XSS) (CSP). Yet, the Google poll indicates that less than 3.7% of actual websites have CSP. The difficulties of deployment and incompatibility with cutting-edge browsers are to blame for CSP's low scalability. In this paper, we propose JavaScript-based CSP (JSCSP), which can handle the majority of real-world browsers and automatically build security policies, in order to investigate the scalability of CSP. In particular, JSCSP provides a revolutionary self-defined security policy that enforces crucial restraints to connected objects, such as JavaScript functions, DOM components, and data access. As opposed to the functionality offered by CSP, JSCSP provides an effective algorithm to automatically produce the policy directives and enforce them in a cascade manner. Our evaluation reveals that the Chrome extension is compatible with well-known JavaScript libraries as a result of the additional implementation of JSCSP. Our JSCSP plugin is capable of identifying and thwarting tried-and-true attack routes taken from popular online applications. We claim that JSCSP outperforms other XSS protection options in terms of performance.

Index Terms—Cross-site Scripting Attacks; Content Security Policy; Origin Confinement; JavaScript Sandbox; Cookie Protection.

1. INTRODUCTION

Due to providing compatibility and friendly user inter-face for modern cloud applications, web services are widely used in practical sectors, such as finance, govern-ment council, and industry. But the security of the services have attracted considerable attention nowadays, because potential vulnerability maybe exploited by attackers to yield severe influence to service providers but also sub- scribers.

Cross-Site Scripting

XSS is one of the most prevalent types of web vulnerability and consistently resides on the OWASP top 10 vulnerability list [1]. If a web page is XXS vulnerable, attackers could inject malicious JavaScript into it and further lead web users to trigger the code execution, so that the sensitive information of the users, which are stored in cookies, session ID and credentials, could be compromised.

Although XSS seems not to be as harmful as other web vulnerabilities, such as SQL injection and code execution,

- G. Xu, X. Xie and S. Huang are with the Tianjin Key Laboratory of Advanced Networking (TANK), College of Intelligence and Comput- ing, Tianjin University, Tianjin, China. Email:losin@tju.edu.cn, xiexi- aofei@tju.edu. cn, 541395961@qq.com
- J. Zhang is with the School of Software and Electrical Engineering, Swinburne University of Technology, Hawthorn, VIC, Australia. Email: junzhang@swin.edu.au
- L. Panis with the School of Information Technology, Deakin University, Geelong, VIC, Australia. Email:l.pan@deakin.edu.au
- W. Lou is with the Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong. Email: csweilou@comp.polyu.edu.hk
- K. Liang is with the Delft university of technology, Netherlands. Email: k.liang-3@tudelft.nl
- Xiaofei Xie and Shuhan Huang are the corresponding authors (Email: xiexi- aofei@tju.edu.cn, 541395961@qq.com).

it is extremely hard to be defended on user side. It is of great importance to design defense mechanisms against XSS attacks to protect end users from losing their credentials, but also to reduce the potential harm yield by worms and malware which are implanted into web page via XSS vulnerability. In practice, there are three main categories of XSS:

- **Reflected XSS.** An attacker injects browser executable code within URI or HTTP parameters. The injection is not stored within the application. Instead, it is non-persistent and only harms users who click the maliciously crafted link to redirect to the third-party web page embedded with malicious code.
- **Stored XSS.** It is also known as persistent XSS. A mali- cious script is injected directly into a web application with a backend server. The script is stored in the server so that any user who visits the application will be harmed.
- **DOM based XSS.** It is an XSS attack modifying the DOM (Document Object Model) in user browser where the original script on user side will be executed in the manner different to its original intension. In addition, there have been other variants of XSS in the literature, such as Mutation XSS (mXSS) [2] , and

Universal XSS (UXSS) [3].

Content Security Policy
Content Security Policy (CSP) [4] is an added security layer that helps detect and mitigate certain types of attack

CSP could be easily bypassed by many approaches. For example, script gadgets (legitimate JavaScript fragments within an application legitimate code base) could be used to execute JavaScript bypassing

CSP directives may bring negative influence on normal _____ functions provided by web applications. For instance,

if
<html>
we add a directive "script-src 'self'" to the CSP header
<script>
like Listing 1, all inline scripts in the web page will be ↪
src='https://code.jquery.com/jquery.min.js'>
disabled, including the normal scripts.
</script>

2. RELATED WORK

In this section, we first introduce works in CSP bypasses and defenses against XSS attacks and further present works about automatic enforcement of CSP.

CSP bypasses. Lekies et al. [7] put forward Code-Reuse Attacks via Script Gadgets, which could bypass CSP. In this attack, the attacker abused so called script gadgets (legiti- mate JavaScript fragments within an application's legitimate code base) to execute JavaScript. Weichselbaum et al. [8]

It refers to pages without malicious scripts. we mark script tags' positions in it and any other scripts outside of these positions will be removed later.

It is similar to the 'nonce' directives in CSP, which only allow the execution of scripts with the right 'nonce' attributes.

Round that 75.81% of distinct policies use script whitelists that allowed attackers to bypass CSP. Moreover, Calzavara et al. [9] investigated the use and effectiveness of CSP as a security mechanism for websites against XSS attacks. They found that existing policies exhibit a number of weaknesses and misconfiguration errors, which might be exploited by attackers to bypass the defense. Doli re et al. [10] found a divergence among browsers implementations in the en- forcement of CSP in `srcdoc` sandboxed `iframes`. Specifi-cally, Gecko-based browsers were proven to have a problem in CSP implementation, which might cause security issues. **Client-side defenses against XSS attacks.** There are works on client-side defenses against XSS attacks. The first policy-based approach on client-side was JSAgents Library [11]. It supported the basic features of CSP 1.1, but it could

not enforce confinements to DOM elements generated dy-namicallyor generate security policies automatically. Panet al. [12] proposed a DOM-XSS detecting framework using static analysis and dynamic symbolic execution. Lekies et al. [13] focused on detecting DOM-based XSS vulnerabilities using ataint analysis approach. DexterJS [14] , [15] was another DOM-based XSS detecting tool, which leverages source-to-source rewriting to carry out character-precise taint tracking when executing in the browser context. In this way, 820 distinct zero-day DOM-XSS attacks were found in Alexa' s top 1000 sites.

Server-side defenses against XSS attacks. There were many server-side solutions to XSS defenses. ModSecurity[7] is an open-source Web Application Firewall, commonly used with the OWASP Core Rule Set. Thome et al. [19] proposed a search-driven constraint solving technique and implemented it as an XSS detection tool. WebMTD [20] randomized certain attributes of DOM elements before de- livering it to the client. Since it was difficult for the attackers to guess the random mapping, the client could distinguish between trusted content and malicious scripts easily. Jin et al. [21] implemented a prototype called NoInjection as a patch to PhoneGap in Android to defend against the XSS attacks in HTML5-based mobile apps. Mohammadi et al. [22] introduced a technique to automatically extract the sanitization functions and then evaluated their effectiveness against attacks using automatically generated attack vectors. Cao et al. [23] proposed PathCutter as a new approach to severing the self-propagation path of JavaScript worms. In addition, some solutions are programming language spe-

https://github.com/cure53/DOMPurify.
https://github.com/google/closure-library.
https://noscript.net/.
http://www.modsecurity.org/.

Automatic generation of CSP. deDacota [26] used a novel approach to secure legacy web applications by automat- ically and statically rewriting an application so that the code and data were clearly separated in its webpages. The separation of code and data could be efficiently enforced at run time via the CSP enforcement mechanism available in modern browsers. CSPAutoGen [27] trained templates for each domain, generated CSPs based on the templates, rewrote incoming webpages on the fly to apply those gen- erated CSPs.

Although there are many other approaches to defend against XSS, CSP is the only standard recognized by the W3C group. Moreover, it could work properly together with many other solutions such as the XSS filter. However, standard CSP relies on the kernel support of browsers and has limited directives. We identify this gap

and address it while designing JSCSP. Thus, JSCSP supports more features such as a cascaded policy language and DOM protection that are not in CSP. Moreover, all of its functional logic are implemented in JavaScript, which greatly extends the list of supported browsers.

3. OVERVIEW OF JSCSP

In this section, we will first introduce the overview of JSCSP as well as the objectives of the design (Section 3 . 1). Then we will describe the fine-grained policy generation (Section 3 . 2) . Next we will use an example to show the format of the generated policy (Section 3.3). Finally, we will discuss the advantage of JSCSP on defending CSP-Bypass attacks (Section 3 . 5) , and the limitation, usage and security of JSCSP (Section 3.6).

Objectives of JSCSP

Fig. 1 shows the overview of the JSCSP, which is designed for addressing the challenges of CSP (see Section 1 . 2) . In general, JSCSP consits of three steps: DOM analysis, auto- matic policy generation and policy enforcement.

In particular, to mitigate the *compatibility* problem of CSP, JSCSP is implemented with JavaScript that supports almost all of browsers. For the *scalability* problem, we design JSCSP with the automatic policy generation and enforcement. Not like CSP that depends on the manual setting on the source code of web applications, JSCSP can automatically generate the policy based on the analysis of the content of the web page. The automation improves the usability of JSCSP significantly. Even if not an expert, the administrator can generate the policy (

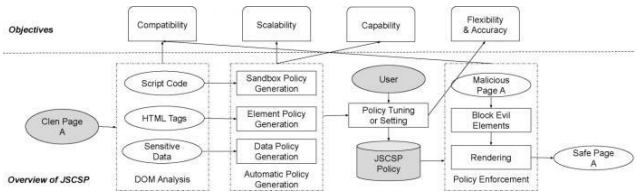


Fig. 1: Overview of the JSCSP Architecture.

will be protected. These policies can restrict the access to the important data from potential malicious JavaScript code.

4. POLICY TUNING OR SETTING

Policy Generation

Before the generation of policy, users should make sure that the current page has not been injected with malicious codes. And new online pages meet this condition. Our policies are generated by analyzing a safe page and enforced in pages that may have been attacked. But users can also design their own policies and further store them into the local Storage.

```
this.execute = function (code) {
var script = JSCSP.doc.createElement('script');
script.setAttribute("class","jscsp-hook");
var code = JSCSP.doc.createTextNode(code);
script.appendChild(code);
JSCSP.doc.head.insertBefore(script, JSCSP.doc.head.children[0]);
}
this . Sandbox_string = function (func_name)
```

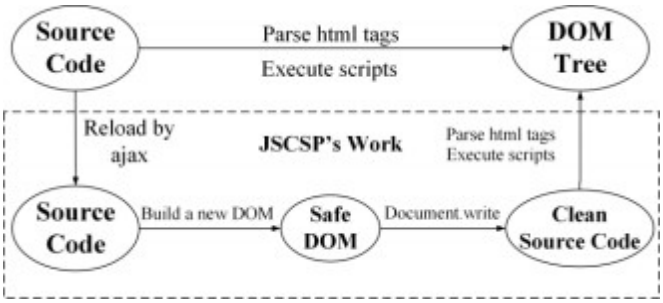
Algorithm 1 The Policy Generation Process

```
// Prepare for policy generation
for all func such that func = function blacklist do
    addFuncHook();
end for
for all data such that data = data list do
    addDataReadHook();
    addDataWriteHook();
end for
    GetScriptPositions();
    GetEventHandlerPositions();
    GetJavaScriptURIPositions();
    GetDataURIPositions();
// Page loaded
setSandboxPolicy(localstorage['sandbox']);
setDataPolicy(localstorage['data']);
elements = document.querySelectorAll('*')
for i = 0; i < elements.length; i + + do
    tagname = elements[i].tagName
    insertSrcWhitelist(tagname, elements[i].src)
end for
for iin dangerous_tag do
    if document.querySelectorAll(dangerous tag[i])
then
    insertTagBlacklist(dangerous tag[i])
    end if
end for
```

Monitor the dangerous function calls. First, we initialize a blacklist which contains vulnerable functions and objects. In order to find whether they are used in the clean page, we add hooks to their *get*[10] method. We embed hooks to the dangerous functions (include objects' constructor such as *Proxy()*). If such a dangerous function

is called, we delete it from the blacklist. In this way, we can get a sandbox policy which forbids all dangerous function calls not used in the clean page. Note that our hook codes must be converted to a string (Listing 9) before the string is injected into the original page, This is so because the code in our extension is in a different context from the original page.

Monitor important data's reading/writing. By default, we use policies to protect important data from malicious access. JSCSP generates data protection policies in a man- ner similar to the previous policy generation process. The only difference is that we use the ES5 functionality of *Ob- ject.defineProperty()* instead of function hooks. This method allows a precise addition to our modification to the property of an object. Accordingly, we are able to use it to modify important data's get or set property (e.g., document.cookie), and to tell if it is accessed by the normal functions in the



Policy Enforcement

There are many types of policies in previous phases. As shown in Algorithm 2, different methods are used to enforce these policies.

Algorithm 2 The Policy Enforcement Process

```
// Stop window from rendering
window.stop();
// Reload html content and seal the new DOM
JSCSP.doc = Reload(currentUrl);
JSCSP.doc = seal(JSCSP.doc);
this.seal = function (doc) {
JSCSP.policy = getPolicy(currentUrl);
for (var item in doc) {
if (typeof doc[item] === 'function') {
// Enforce the given Policies.
Object.defineProperty(
for all func such that func = {
item
{ value:
doc[item],
JSCSP.policy[ sandbox ] do
);
configurable:
false }
deleteFunc(func);
}
end for
}
elementPolicies = JSCSP.policy[ 'element ];
return doc;
for all selector such that selector = elementPolicies
}
do
policy = elementPolicies[selector];
for all element such that
element =
JSCSP.doc.querySelectorAll(selector) do
checkElement(element, policy);
flagElement( element);
end for
end for
checkCodeReuseV ectors();
checkScriptPositions();
checkEventHandlerPositions();
checkJavaScriptURIPositions();
checkDataURIPositions();
for all element such that element = elementFlagged
do
if element.allow = false then
deleteElement( element);
end
if
end for
// Copy back to origin DOM
startDocument(JSCSP.doc.documentElement.inn _
erHTML);

// Restrict dynamic elements.
hookFunctions();
// Data protection.
hookData();
```

Fig. 4: The way of loading pages when JSCSP is enabled.

Stop the window and reload the page. Before JSCSP imposes confinements on the DOM according to the policies, the window object should be stopped as early as possible. It is necessary to prevent malicious scripts' execution and win possible attacker-caused race-conditions (such as DOM- clobbering [30]). After that, we reload the page and get its html code by using XMLHttpRequest (Fig. 4). We now extract the html contents and map them into a safe DOM11.

Listing 11: Seal the safe DOM

Seal the safe DOM. In case of the loss of race-conditions, we need to make sure that the existing DOM properties and built-in functions have not been tampered with. Thus JSCSPiterates overall methods to lock them to prevent from external accesses (Listing 11).

Enforce the given policies. There are three types of en-forcers in JSCSP, which are designed for the corresponding security policies respectively.

Sandbox policies enforcer. It deletes the methods that are forbidden from their owner by using the delete statement. In fact, we cannot delete the methods of the window object directly. Because our injected scripts run in the context of each individual web page, only DOM elements can be modified. Thus, additional scripts are injected into the context of the original web page using inline scripts.

Element policies enforcer. We use the document query. Selector All API to select all items in the DOM tree. The enforcer requests all elements matching the JSCSP policy selectors and passes them to the corresponding enforcer methods. After the final selector's rules have been enforced, the elements are actually removed or modified (see the next subsection). Moreover, in order to defend against the code-reuse attacks, we add additional rules[12] to filter DOM elements.

UXSS attacks, we chose 4 UXSS vulnerabilities. Attackers could use them to bypass CSP confinements. However, JSCSP was not affected and could still defend XSS attacks properly. In addition, redirection attacks (e.g., win- dow.location="http://attacker. com/ c=" +document. cookie) were blocked by JSCSP. In Code-Reuse attacks' experiments, we found that POC vectors could be inserted when specific Script Gadgets were used (e.g., Vue.js) because the attack vectors of these gadgets were similar to the normal

It is an offensive security's exploit database archive. https:// www.exploit-db.com/.

eral rare HTML5 XSS vectors (e.g., <iframe srcdoc "<svg onload = alert(1)>"></iframe>) can also be defended against by JSCSP. The detailed result is listed in Table 2, which shows that though JSCSP is bypassed by much less of POC vectors, all exploit vectors are prevented from stealing cookie successfully.14. https://link.springer.com/chapter/10.1007/978-3-319-24174.

5. CONCLUSION AND FUTURE WORK

JSCSP isan exciting development in the ongoing effort to improve web security. By providing a flexible and powerful solution for mitigating cross-site scripting attacks, JSCSP has the potential to make the web a safer place for everyone. Further research and development are needed to fully realize the potential of JSCSP and to address other web-based security threats.

6. ACKNOWLEDGMENTS

This work is partially sponsored by the State key development program of China (No. 2018YFB0804402, 2019YFB2101700), National Science Foundation of China (U1736115), and Hong Kong Polytechnic University under Grants (BCB6, YBJU, UAH6, UAJH).

7. REFERENCES

[1] The MITRE Corporation. *Common Vulnerabilities and Exposures: The Standard for Information Security Vulnerability Names* [EB/OL]. 2017[2017-09- 29]. http://cve.mitre.org.

[2] Heiderich, J. Schwenk, T. Frosch, J. Magazinius, and E.Z. Yang, "*mxss attacks: attacking well-secured web-applications by using innerhtml mutations*," in *CCS*, Berlin, Germany, 2013, pp. 777-788.

[3] The Acunetix. Universal Cross-site Scripting (UXSS): The Making of a Vulnerability [EB/OL]. 2017[2017- 10- 2]. https://www.acunetix.com/blog/articles/universal-cross-site-scripting-uxss.

[4] S. Stamm, B. Sterne, and G. Markham, "Reining in the web with content security policy," in *WWW*, Raleigh, 2010, pp. 921-930.

[5] The W3C Working Draft. *Content Security Policy Level 3*[EB/OL]. 2017[2017 -09-29]. https://www.w3.org/TR/CSP3/.

❑❑❑

Multi-authority Keyword Search Using Attributes Over Encrypted Cloud Data

T.P. UDHAYASANKAR[1], M. KIRUBA[2], K.MOHANA PRIYA[3], D.VASANTH[4]
[1]Associate Professor, [2,3,4]Student
[1,2,3,4]Department of CSE, Annapoorana Engineering College, Salem, Tamil Nadu, India

ABSTRACT

A crucial method for ensuring both data security and accessibility in the cloud is searchable encryption (SE). The Cipher text-Policy Attribute-Based Keyword Search(CP-ABKS) approach can concurrently achieve keyword-based retrieval and fine grained access control by utilizing Cipher text-Policy Attribute-Based Encryption (CP- ABE). Yet, in current CP-ABKS schemes, the secret key distribution and pricey user certificate verification are the sole responsibilities of the single attribute authority. Moreover, this causesdistributed cloud systems to experience a single-point performance bottleneck. Due to these restrictions, we provide a secure Multi-authority CP-ABKS (MABKS) system in this study to solve them and reduce the computational and storage loadon devices with restricted resources in cloud systems. The MABKS system has also been expanded to include features for malicious attribute authority tracing and attribute updating. The MABKS system is selectively secure in both selective-matrix and selective-attribute models, according to our in-depth security research. Our experimental findings utilizing actual datasets show the effectiveness and usefulness of the MABKS system in real-world settings.

Index Terms— Access Control, Cloud Storage, Multiauthority Ciphertext-Policy Attribute-Based Encryption (CP-ABE), Public Update, Revocation.

1. INTRODUCTION

Now a day's cloud computing is an intelligently developed technology to store data from number of clients. Cloud computing allows users to remotely store their data over cloud. Remote backup system is the progressive technique which minimizes the cost of implementing more memory in an organization. It helps government agenciesand enterprises to reduce financial overhead of data management. They can extract their data backups remotely to third party cloud storage providers than maintaining their own data centers. An individual or an organization does not require purchasing the storage devices. Instead, they can store their data to the cloud and archive data to avoid information loss in case of system failure like hardware or software failures. Cloud storage is more flexible, but security and privacy are available for the outsourced data becomes a serious concern.

Access control methods ensure that authorized user access data of the system. Accesscontrol is a policy or procedure that allows, denies or restricts access to system. It also monitors and record all attempts made to access a system. Access Control can also identify unauthorized users attempting to access a system. It is a mechanism which is very much important for protection in computer security. The Cloud storage is a very important service in cloud computing. The Cloud Storage offers services for data owners to host their data over cloud environment. longer rely on servers to do access control, so the data access control becomes a challenging issue in cloud storage systems. Therefore, the decentralized data access control scheme is introduced.

2. LITERATURE SURVEY

DAC-MACS: Effective data access control for multi-authority cloud storage s ystems

Data access control is an effective way to ensure the data security in the cloud. However, due to data outsourcing and untrusted cloud servers, the data access control becomes a challenging issue in cloud storage systems. Existing access control schemesare no longer applicable to cloud storage systems, because they either produce multiple encrypted copies of the same data or require a fully trusted cloud server. Ciphertext-Policy Attribute-based Encryption (CP-ABE) is a promising technique for access controlof encrypted data.

DACC: Distributed Access Control in Clouds

We propose a new model for data storage and access in clouds. Our scheme avoids storing multiple encrypted copies of same data. In our framework for secure data storage, cloud stores encrypted data (without being able to decrypt them). The main novelty of our model is addition of key distribution centers (KDCs). We propose DACC (Distributed Access Control in Clouds) algorithm, where one or more KDCs distribute keys to data owners and users. KDC may provide access to particular fields in all records. The scheme is collusion secure; two users cannot together decode any data that none of them has individual right to access. DACC alsosupports revocation of users, without redistributing keys to all the users of cloud services. We show that our approach results in lower communication, computation and storage overheads, compared to existing models and schemes.

Expressive, efficient and revocable data access control for multi-authority cloud storage

Data access control is an effective way to ensure the data security in the cloud. Due todata outsourcing and untrusted cloud servers, the data access control becomes a challenging issue in cloud storage systems. Ciphertext-Policy Attribute-based Encryption (CP-ABE) is regarded as one of the most suitable technologies for data access control in cloud storage, because it gives data owners more direct control on access policies.

3. PROPOSED SYSTEM

Architecture with many tiers of authority. For the first time, a hierarchical structure in the MABKS system enables many AAs to carry out the time-consuming user certificate verification and intermediate secret key generation on behalf of the CA, reducing the CA's computation requirements. Keywords can be searched at the file level. MABKS differs from standard CP-ABKS methods in that the secret key used toencrypt a file's file key is embedded within the indexing process, rather than separateoperations. [4], [5], [12] For cloud clients (such as data owners and users), this meansthat the MABKS system allows them to execute keyword-based ciphertext retrieval, as well as file-level fine-grained encryption access control.

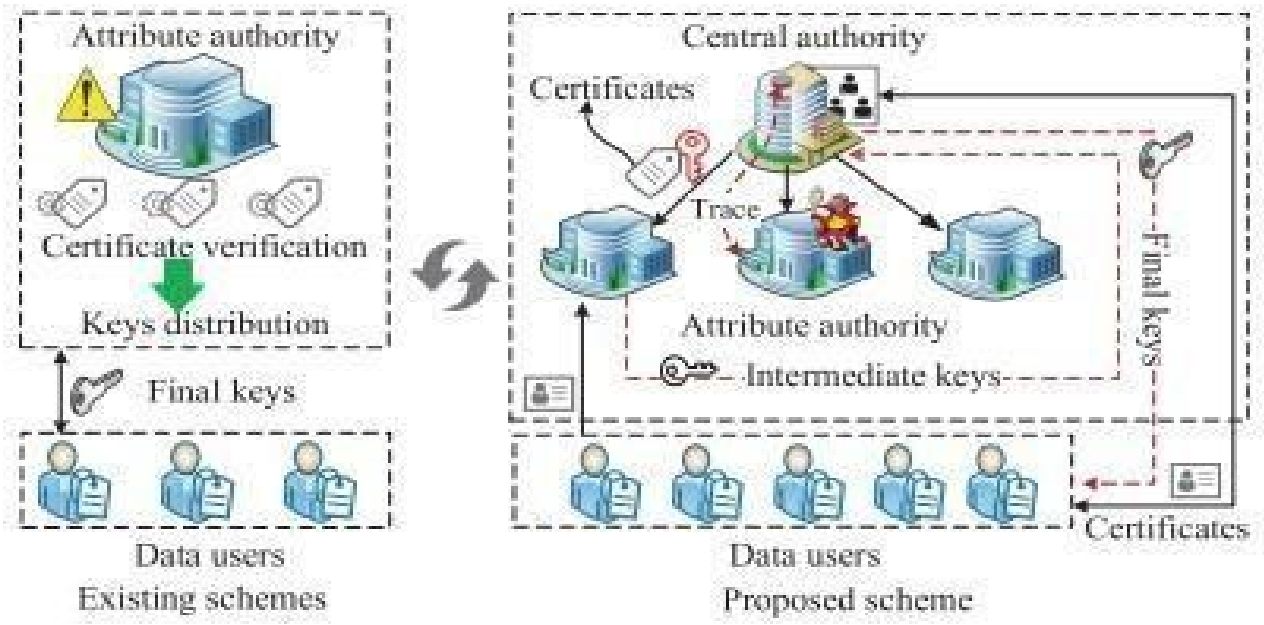


Fig 1: Architecture

AA (Attribute Authority) It is responsible for user legitimacy verification procedure, and then sending an intermediate key to CA for legitimacy verified users. AAs can work simultaneously to perform user legitimacy verification. When any useraccesses any type of data, AA informs the owner of respective data by a message containing the username.

Fig 2: Home Page

Fig 3: Uploaded File Details

Fig 4: User Request Page

4. CONCLUSION

A multi-authority keyword search scheme using attributes over encrypted cloud data. The scheme enables multiple authorities to define access policies based on attributes of the data, such as the data owner, data type, and data creation date. The scheme ensures that only authorized users can search and retrieve the encrypted data from the cloud while preserving the confidentiality of the data. We conducted a security analysis and experimental validation of the scheme and showed that it satisfies the confidentiality, integrity, and availability requirements and is efficient in terms of search time and communication overhead. Our proposed scheme provides a practical solution for secure and efficient keyword search over encrypted cloud data.

5. REFERENCES

[1] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in Proc. Adv. *Cryptol.— EUROCRYPT 2005*. New York, NY, USA: Springer, 2005, pp. 457–473.

[2] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption forfine-grainedaccess control of encrypted data," in *Proc. 13th ACM Conf. Comput. Commun. Security*, 2006, pp. 89–98.

[3] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in *Proc. IEEE Security Privacy 2007*, 2007, pp. 321–334.

[4] M. Pirretti, P. Traynor, P. McDaniel, and B. Waters, "Secure attribute-based systems," in Proc. 13th ACM Conf. Comput. Commun. Security, 2006, pp. 99–112.

[5] S. Yu, C. Wang, K. Ren, and W. Lou, "Attribute based data sharing with attribute revocation," in *Proc. 5th ACM Symp. Inf., Comput. Commun. Security 2010*,2010, pp. 261–270.

[6] S. S. M. Chow, "A framework of multi-authority attribute-based encryp-tion with outsourcing and revocation," in *Proc. 21st ACM Symp. Access Control ModelsTechnol.*, 2016, pp. 215–226.

[7] J. Hur and D. K. Noh, "Attribute-based access control with efficient revo- cationin data outsourcing systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 7, pp. 1214–1221, Jul. 2011.

[8] C.-I. Fan, V. S.-M. Huang, and H.-M. Ruan, "Arbitrary-state attribute-based encryption with dynamic membership," *IEEE Trans. Comput.*, vol. 63, no. 8, pp. 1951–1961, Aug. 2014.

[9] K. Yang, X. Jia, K. Ren, B. Zhang, and R. Xie, "DAC-MACS: Effective data access control for multiauthority cloud storage systems," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 11, pp. 1790–1801, Nov. 2013.

[10] S. Ruj, A. Nayak, and I. Stojmenovic, "DACC: Distributed access control in clouds," in *Proc. 2011 IEEE 10th Int. Conf. Trust, Security Privacy Comput. Commun.*, 2011, pp. 91–98.

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Breast Cancer Detection Using Pre-Processed Images

SMT.S J R K PADMINIVALLI V[1], VALLURU KOMALI[2], NELLURI NAGA SAI KRISHNA[3], KUNCHALA SUMANTH[4]
[1]Assistant Professor, [2,3,4]Student
[1,2,3,4]Department of CSE, RVR & JC College of Engineering, Guntur, Andhra Pradesh, India

ABSTRACT

One of the leading causes of cancer death in women is breast cancer. It occurs when cells in the breast begin to grow uncontrollably, often forming a lump or mass. Although it can happen to both men and women, it affects women more frequently. There are many Machine Learning methods available to detect this cancer but they provide less accuracy and this can potentially be resolved by implementing effective image pre-processing techniques like background removal, noise reduction and image enhancements. The pre-processed images are then sent to models like Convolutional Neural Network (CNN), Decision Tree and K-Nearest Neighbor and the results are combined using various techniques to provide better accuracy.

Index Terms— Pre-Processing; CNN; KNN; Decision Tree.

1. INTRODUCTION

Breast cancer is the second most common type of cancer among women worldwide, with an estimated 2.3 million new cases diagnosed in 2020 alone. Breast cancer can occur at any age, but it is most common in women over the age of 50. There are several types of breast cancer, including ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), and invasive lobular carcinoma (ILC).Risk factors for breast cancer include being female, older age, a personal or family history of breast cancer, certain gene mutations (such as BRCA1 and BRCA2), dense breast tissue, and exposure to estrogen. Hence it is important to increase awareness about breast cancer and encourage women to have a regular checkup and save their lives.

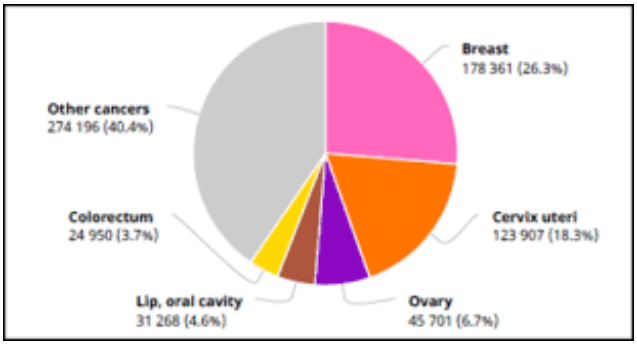


Figure 1: Cancer Statistics

Mammography is considered the gold standard for regular screening. This screened data are to be analyzed which requires radiologists but, the shortage of radiologists leads to delay in treatment. Hence ,it is important to develop an intelligent system that can detect and diagnose abnormalities quickly and accurately.Before developing an intelligent system, it is important to effectively pre-process mammographic images. This involves removing the background, pectoral muscle, and the addition of noise along with the application of image enhancements.

2. RELATED WORK

Detecting cancer from mammographic images is a difficult, hence important features from these images should be iden- tified. Several approaches exist to detect and segment masses in mammographic images[1] automatically, and the key points and main differences between those strategies are highlighted. The main objective is to highlight the advantages and disadvantages of the different approaches. This review provides a quantitative comparison in addition to the qualitative description and com- parison of different approaches. Two mammographic databases are compared to compare the performance of seven mass de- tection methods. One is a public digitised database, while the other is a local full-field digital database for local detection. A Receiver Operating Characteristic (ROC) and a Free-response Receiver Operating Characteristic (FROC) analysis is presented. Since the American Cancer Society (ACS) issued recommen- dations for early detection of breast cancer in 2003, new infor- mation regarding breast magnetic resonance imaging (MRI)[5] screening has become available. A guideline panel evaluated the study and made new recommendations for women at various risk levels. A screening MRI is recommended for women with a lifetime risk of breast cancer greater than or equal to 20-25% of hers. This includes women with a significant family history of breast cancer and women who have been treated for Hodgkin's disease. There is insufficient evidence to support screening in women with a history of breast cancer, women with cancer in situ, women with atypical hyperplasia, and women with very dense breasts on mammography. risk subgroup. The review was assumed to be out of scope including the diagnostic use of MRI.

Radiologists prefer to employ breast ultrasound and

mamography imaging modalities to visualize breast cancer. A area of interest (ROI)[11] indicating the tumor is extracted from the image in order to find malignancy. When noise, low contrast, and blurriness are present, segmentation becomes laborious. Be- fore to segmentation, pre-processing is done to improve contrast and remove extraneous information from the image. The catego- rization of the picture into benign and malignant classifications is also influenced by segmentation. The literature has suggested a number of segmentation approaches to separate the pectoral muscles from the microcalcification region of interest, masses, and breast lesions. This paper offers a thorough analysis of various methods, especially as they pertain to mammography pictures.

Machine Learning algorithms are also applied on datasets of UCI repository for breast cancer detection[3],the work intro- duced in this paper objectives to analyse the overall performance of a couple of computing device gaining knowledge of classifiers in detecting breast cancer. The acquired consequences genuinely point out the efficacy of linear guide vector classifier and gra- dient boosting over different classifiers. The findings of this lookup can be similarly utilized for creating extra environment friendly ensemble fashions as properly as optimizing the overall performance of present fashions thereby growing their predic- tion accuracy.

Later, Computer Aided Diagnosis(CAD) were used to as- sist radiologists in their work by carrying out a double-reading procedure that offers a second opinion that the doctor might con- sider during the detection phase.[12] study presents a computer-aided design (CAD) model for the detection of suspicious re- gions that are later classified as benign or malignant based on a set of features extracted from lesions to describe their visual content by support vector machines, artificial neural networks, and linear discriminant analysis. To identify the subset of fea- tures with the highest discriminant power, a genetic algorithm is applied.

Early breast cancer identification and other abnormalities in human breast tissue are now possible thanks to digital mammog- raphy. It gives us the chance to create algorithms for computer- aided detection (CAD).3 seperate steps were suggested in [8]. The first stage entails improving the contrast using the contrast limited adaptive histogram equalization (CLAHE) method. Af- ter that, create the rectangle to separate the pectoral muscle from the region of interest (ROI), and then use our suggested mod- ified seeded region growing (SRG) technique to suppress the pectoral muscle. The 322 mammography pictures in the MIAS database were all subjected to the proposed algorithms, which resulted in total pectoral muscle suppression in the majority of the scans. As compared to previous segmentation approaches, the suggested algorithm produces better results.

One of the illnesses that causes a significant number of fa- talities each year is breast cancer. In the entire world, it is the most prevalent type of cancer and the leading cause of mortality for women. Particularly in the medical industry, where those techniques are frequently utilized to make judgments through diagnosis and analysis. [6] compares the performance of four machine learning algorithms on the datasets: Support Vector Machine (SVM), Decision Tree (C4.5), Naive Bayes (NB), and k Nearest Neighbors (k-NN). The major goal is to evaluate each algorithm's efficiency and efficacy in terms of accuracy, preci- sion, sensitivity, and specificity in order to determine whether or not the data classification was right. According to experimental findings, SVM provides the highest accuracy and lowest error rate.

3. DATASET

Mammography is a medical imaging technique that uses low- dose X-rays to examine breast tissue for signs of abnormalities or cancer. It is the most common screening tool used for breast cancer detection. The American Cancer Society recommends that women with an average risk of breast cancer begin annual mammograms at age 45, and then switch to screening every other year at age 55. Women at higher risk of breast cancer may need to start screening earlier and/or have more frequent mam- mograms. Mammogram images from "cbis-ddsm-breast-cancer- image-dataset" are used to implement the proposed methods. This CBIS-DDSM is a modified and standardized version of DDSM (Digital Database for Screening Mammography). The image is a jpeg of the original dataset (163GB). Original reso- lution was maintained.The dataset contains 38 attributes like filepath,image-path,BitsAllocated,BitsStored.The dataset con- tains about 10239 images.It contains normal, benign, and malig- nant cases with verified pathology information.

Column Non-Null Count

0 file_path 10237 non-null
1 image_path 10237 non-null 2 AccessionNumber 0 non-null
3 BitsAllocated 10237 non-null
4 BitsStored 10237 non-null
5 BodyPartExamined 10237 non-null
6 Columns 10237 non-null
7 ContentDate 10237 non-null
8 ContentTime 10237 non-null
9 ConversionType 10237 non-null
10 HighBit 10237 non-null
11 InstanceNumber 10237 non-null
12 LargestImagePixelValue 10237 non-null
13 Laterality 9671 non-null
14 Modality 10237 non-null
15 PatientBirthDate 0 non-null
16 PatientID 10237 non-null
17 PatientName 10237 non-null
18 PatientOrientation 10237 non-null
19 PatientSex 0 non-null 20 PhotometricInterpretation 10237 non-null
21 PixelRepresentation 10237 non-null 22 ReferringPhysicianName 0 non-null
23 Rows 10237 non-null
24 SOPClassUID 10237 non-null
25 SOPInstanceUID 10237 non-null 26 SamplesPerPixel 10237 non-null
27 SecondaryCaptureDeviceManufacturer 10237 non-null
28 SecondaryCaptureDeviceManufacturerModelName 10237 non-null
29 SeriesDescription 9671 non-null
30 SeriesInstanceUID 10237 non-null
31 SeriesNumber 10237 non-null
32 SmallestImagePixelValue 10237 non-null
33 SpecificCharacterSet 10237 non-null
34 StudyDate 9671 non-null
35 StudyID 10237 non-null
36 StudyInstanceUID 10237 non-null
37 StudyTime 9671 non-null

Figure 2: Dataset Description

4. METHODOLOGY

Image Pre-Processing

Image preprocessing is an essential step in image analysis and computer vision tasks. Preprocessing techniques like noise re- duction, image enhancement, image segmentation, normaliza- tion are applied on the images. Noise is an unwanted and ran- dom variation in the pixel values of an image, which can be introduced during image acquisition or transmission. Noise reduction is applied to improve visual quality, enhance image analysis, increase signal-to-noise ratio improving compression. Image enhancement techniques can be used to improve the con- trast, image sharpening, making the image clearer and easier to analyze. Normalization is a crucial step in image preprocess- ing that involves transforming the pixel values of an image to a common scale. It can be used for improved training, consis- tent performance ,improve visualization, reduce computational complexity.

Convolutional Neural Networks

CNNs[14] are a type of deep learning model that can automat- ically learn to recognize complex patterns and features in im- ages.Breast cancer detection typically involves analyzing medi- cal images, such as mammograms or MRI scans. CNNs[15],[10] can be trained on large datasets of labeled medical images to accurately identify potential cancerous regions and distinguish them from normal tissue.One approach is to use a CNN as a bi- nary classifier, where the model is trained to predict whether an image contains cancerous tissue or not. Another approach is to use a CNN as a segmentation model, where the model is trained to identify the exact location and extent of cancerous regions within an image, here we use it as a classifier model.Several studies have shown that CNNs can achieve high accuracy in breast cancer detection and diagnosis.CNN contains convolution layer,max pool layer and dense layers.For breast cancer detection 4 convolutional layers are used with input pattern of (50,50,3) and filters of size 32,64,128 are used.Activation functions[4] like softmax[9],relu[7] are used.

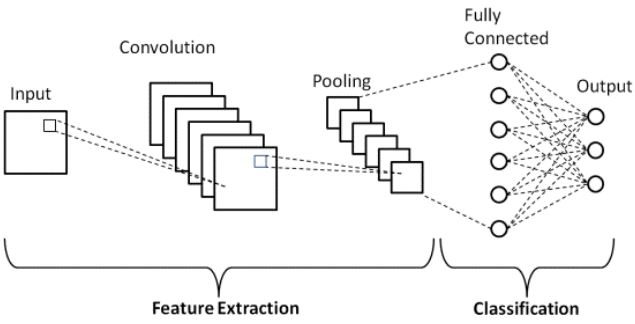


Figure 3: Architecture Of CNN

KNN

K-nearest neighbors (KNN)[13] algorithm is a simple, yet ef- fective method for classification tasks that works well in many situations, including breast cancer classification. KNN is a non- parametric algorithm that does not make any assumptions about the underlying distribution of the data. Instead, it relies on the similarity between the new instance (in this case, a breast cancer case) and the existing labeled data to make a prediction.

In the case of breast cancer classification, the KNN algorithm can be used to classify a new case as malignant or benign based on its similarity to previously diagnosed cases. Specifically, KNN calculates the distance between the features of the new case and the features of the training data. It then selects the k-nearest neighbors based on the smallest distances and assigns the class label of the majority of those neighbors to the new case.

One advantage of using KNN for breast cancer classification is that it does not require a priori knowledge of the underlying probability distribution or parameter estimation. Additionally, KNN can handle large datasets and high-dimensional feature spaces, making it suitable for complex classification problems like breast cancer diagnosis.

5. DECISION TREE

Decision trees[2] are commonly used for classification problems, including breast cancer classification. Decision trees are easy to interpret, as they can be represented graphically, with the deci- sion nodes and branches clearly indicating the decision-making

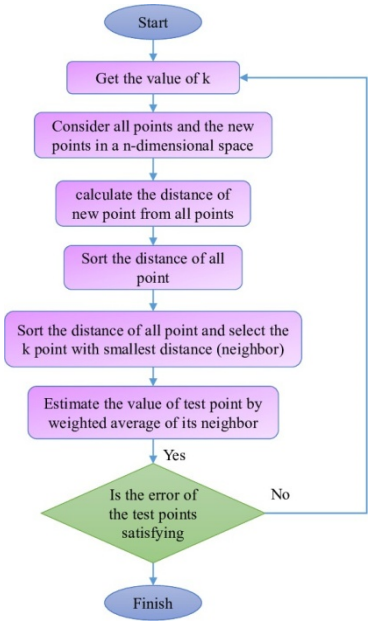


Figure 4: Flowchart of KNN algorithm

process. This is especially important in medical diagnosis, where it is crucial to understand the reasoning behind the decision. De- cision trees can handle mixed data types, such as numerical and categorical data, which is common in medical data. Decision trees can handle missing data by imputing the missing values or creating a separate

branch for missing values, which is impor- tant in medical data where missing data is common. Decision trees can capture complex relationships between features, which is useful in medical diagnosis, where multiple factors can influ- ence the diagnosis.

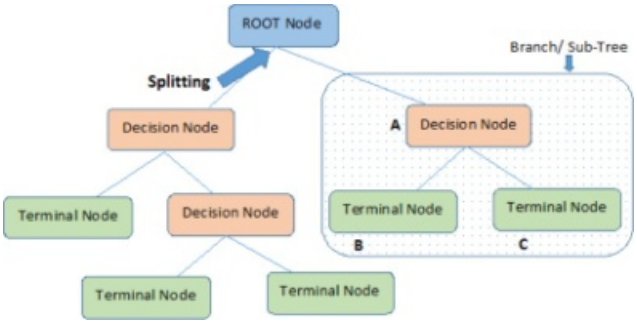


Figure 5: Decision Tree Classification

6. FUSION OF RESULTS

In general, each classifier method, even the same method with different parameters, focuses on different perspectives and em- phasizes certain aspects, and each learning algorithm has its own strengths and weaknesses. Therefore, it is difficult to assess which learning algorithms lead to better performance across all feature sets. The same dataset may be suitable for different algo- rithms to learn effectively. At the measurement level, there are several methods of fusion. Direct join rules are called votes and simply count the votes for each class and return the class with the most votes. Common algebraic combination rules include mean, maximum, and product.

let, $y\ c_1, c_2, \ldots, c_n$, m classifier $h_1, h_2, \ldots, h_m$ are trained in a feature space $D : x_1, x_2, \ldots, x_k$, the combination rules are defined as follows.

$c \qquad x\ , x\ , \ldots, x$

$\underline{1}$ m

P c

7. CONCLUSION

This paper presented the principles of three types of classification methods, KNN, Decision Tree, and CNN, and applied them to breast cancer detection, then improved the performance by

$$P_{\max}(c_i \mid x_1, x_2, \ldots, x_k) = \max_{j=1}^{m} P_{h_j}(c_i^j \mid x_1, x_2, \ldots, x_k)$$

where P_{h_j} is the posterior probability of the classifier h_j, P_{mean} is the posterior probability after averaging fusion for m classi- fiers, $P_{\max}$ is the posterior probability after maximizing for m classifiers, P_{product} is the posterior probability after production for m classifiers.

Firstly, we can obtain the posterior probabilities P_{h_j} of the classifier h_j from the classification results. Once you have two or more sets of classification results, choose an appropriate fusion strategy such as matching, importance, or maximization. Then we can get the predicted posterior probabilities according to the rules of the fusion strategy given in the above equation.

Using simple algebraic rules for classifier fusion can reduce the risk of posterior probability estimation errors and overfitting. Moreover, these algebraic fusion rules have low computationalcomplexity.

8. ANALYSIS OF RESULTS

Initially, a decision tree classifier is built with entropy as the criterion for splitting, a KNN classifier is built by choosing the 3 nearest values, and a CNN classifier is built with specified parameters. After building these classifiers, the models are trained with the same data to find the accuracy of each model. According to Table 1, each model is accurate to a different degree. It is observed that the accuracy of CNN is higher than the other 2 models.

TABLE 1: The performances of DECISION TREE, KNN andCNN.

Classifier	Accuracy
CNN	0.95
KNN	0.90
DECISION TREE	0.85

Fusing results from different classifiers, also known as ensemble learning or model fusion, is a technique used in machine learning and data mining to combine the predictions of multiple classifiers to improve overall prediction accuracy and robustness. Ensemble learning has been shown to improve prediction accuracy compared to individual classifiers. By combining the predictions of multiple classifiers, the ensemble can benefit from the strengths of different classifiers while mitigating their weaknesses. As a result, predictions that are more accurate and reliable can be made. Table 2 shows the result of the model after combining the results of three models. It is quite evident that the accuracy has increased.

TABLE 2: The final fusion results for DECISION TREE,KNN and CNN.

Fusion Method	Accuracy mean
0.97	max 0.96

fusing two perspectives: the fusion of the classification results of KNN, Decision Tree, and CNN using mean and maximum. The results showed that fusion processing is a viable way to improve the performance of breast cancer

detection.

In conclusion, fusing results from different classifiers through ensemble learning can lead to improved prediction accuracy, increased robustness, better handling of model uncertainty, in- creased diversity, better generalization, and increased model stability. Ensemble learning is a powerful technique that can enhance the performance and reliability of prediction models in various machine learning and data mining tasks.

9. REFERENCES

[1] "A review of automatic mass detection and segmentation in mammographic images" by Arnau Oliver, Jordi Freixenet, Joan Martí, Elsa Pérez, Josep Pont, Erika R.E. Denton, Reyer Zwiggelaar.

[2] B. Gambäck and U. K. Sikdar, "Using convolutional neural networks to classify hate-speech," in Proc. 1st Workshop Abusive Lang. Online, 2017, pp. 85–90.

[3] "Breast Cancer Detection Using Supervised Machine Learning: A Comparative Analysis" by Akansha Kamboj, Prash-mit Tanay, Akash Sinha & Prabhat Kumar.

[4] C. Gulcehre, M. Moczulski, M. Denil and Y. Bengio, "Noisy activation functions", Proceedings of International Con-ference on Machine Learning, pp. 3059-3068, 2016.

[5] D. Saslow, C. Boetes, W. Burke, S. Harms, M. O. Leach, C. D. Lehman, E. Morris, E. Pisano, M. Schnall, S. Sener, R. A.Smith, E. Warner, M. Yaffe, K. S. Andrews, and C. A. Russell, "American cancer society guidelines for breast screening with MRI as an adjunct to mammography," CA A, Cancer J. Clinicians, vol. 57, no. 2, pp. 75–89, Mar. 2007.

[6] H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk pre- diction and diagnosis," Pro-cedia Comput. Sci., vol. 83, pp.1064–1069, 2016.

[7] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 2017, pp. 2684-2691, doi: 10.1109/IJCNN.2017.7966185.

[8] I. K. Maitra, S. Nag, and S. K. Bandyopadhyay, "Technique for preprocessing of digital mammogram," Comput. Methods Programs Biomed., vol. 107, no. 2, pp. 175–188, Aug. 2012.

[9] I. Kouretas and V. Paliouras, "Simplified Hardware Im-plementation of the Softmax Activation Function," 2019 8th In- ternational Conference on Modern Circuits and Systems Tech- nologies (MOCAST), Thessaloniki, Greece, 2019, pp. 1-4, doi:10.1109/MOCAST.2019.8741677.

[10] Iztok A. Pilih, Dunja Mladenic, Nada Lavrae and Tine S. Prevec3B. Smith, "Using Machine Learning for Outcome Pre- diction of Patients with Severe Head Injury", Tenth IEEE Sympo- sium on Computer-Based Medical Systems.

[11] J. Dabass, S. Arora, R. Vig, and M. Hanmandlu, "Seg-mentation techniques for breast cancer imaging modalities—A review," in Proc. 9th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence), Jan. 2019, pp. 658–66

[12] J. Suckling, J. Parker, and D. Dance, "The mammographic image analysis society digital mammogram database excerpta medica," in Proc. Int. Congr. Ser., vol. 1069, 1994, pp. 375–378.[13]. M. Sharma and S. Sharma, "Generalized K-Nearest Neighbour Algorithm- A Predicting Tool", Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 3, no. 11, Nov. 2013.

[13] S. Pang, A. Du, M. A. Orgun and Z. Yu, "A novel fused convolutional neural network for biomedical image classifica- tion", Med. Biol. Eng. Comput., vol. 57, no. 1, pp. 107-121,2018.

[14] Y. Kim, "Convolutional neural networks for sentence classification," in Proc. EMNLP, Oct. 2014, pp. 1746–1751.

❑❑❑

VPPCS: Vanet-Based Privacy-Preserving Communication Scheme

MANOJ. M[1], JAYASURYA. J[2], MURALITHARAN. Y[3], SURIYA. S[4]
[1]Assistant Professor, [2,3,4]Student
[1,2,3,4]Department of CSE, AVS Engineering College, Salem, Tamil Nadu, India

ABSTRACT

Over the past years, vehicular ad hoc networks (VANETs) have been commonly used in intelligent traffic systems. VANET's design encompasses critical features that include autonomy, distributed networking, and rapidly changing topology. The characteristics of VANET and its implementations for road safety have attracted considerable industry and academia interest, particularly in research involving transport systems enhancement that could potentially save lives. Message broadcasting in an open access system, such as VANET, is the main and utmost challenging problem with regard to security and privacy in VANETs. Various studies on VANET security and privacy have been proposed. Nevertheless, none has considered overall privacy requirements such as unobservability. In order to address these shortcomings, we propose a VANET based privacy-preserving communication scheme (VPPCS), which meets the requirements for con- tent and contextual privacy. It leverages elliptic curve cryptography (ECC) and an identity-based encryptionscheme. We have carried out a detailed security analysis (burrows–abadi–needham (BAN) logic, random oracle model, security of proof, and security attributes) to validate and verify the proposed scheme. The analysis has shown that our scheme is secure and also shown to be effective in a performance evaluation. The proposed scheme does not only meet the previously mentioned security and privacy requirements, but also impervious to various types of attacks such as replay, impersonation, modification, and man-in-the-middleattacks.

Index Terms— Vanet, PPC Scheme.

1. INTRODUCTION

As the design of wireless communication technology and network systems is continuously and rapidly progressing, vehicular ad hoc networks (VANETs) have regained attention and interest in support of wireless vehicles in communicating with other vehicles and roadside units (RSUs) to guaran- tee traffic safety and improve driving experience [1]–[3]. VANETs also have the benefits of preventing collisions, lane- fusion, optimizing traffic, collecting toll, location-based ser- vices and infotainment [4]–[7]. VANET is basically Mobile ad hoc networks (MANETs) associated with vehicles and RSUs. In contrast to the nodes in a MANET, the power, storage, and computing capacity of vehicles are typicallynot resource constrained. Typical VANET contains trusted authorities (TAs), RSUs (e.g., road-side or other facilities),and onboard units (OBUs) equipped in vehicles [8], [9], as shown in Figure 1.

Using dedicated short-range communication (DSRC) pro- tocol, the communication of VANET can be divided into vehicle-to-vehicle (V2V) and vehicle-to- infrastructure (V2I) [8]. The OBU in the vehicle and the DSRC protocol will allow all vehicles to communicate onthe roadside with adja- cent vehicles and nearby RSUs. For example, traffic related messages on vehicle OBUs regularly broadcasts data on elements such as location, meteorological conditions, route, velocity, and traffic condition. The traffic-related message enables theparticipating vehicles in the region to take the necessary measures to prevent traffic accidents and avoid traf- fic congestion [10]. The traffic-related message (e.g., recenttraffic incidents) may also be forwarded by the RSU and other vehicles to the traffic administration department and other relevant departments (e.g., the traffic police or fire department) to ensure necessary actions can be taken within is inherent in existing VANET schemes. More specifi-cally, the scheme describes the contributions of VPPCS as follows:

2. SECURITY ANALYSIS

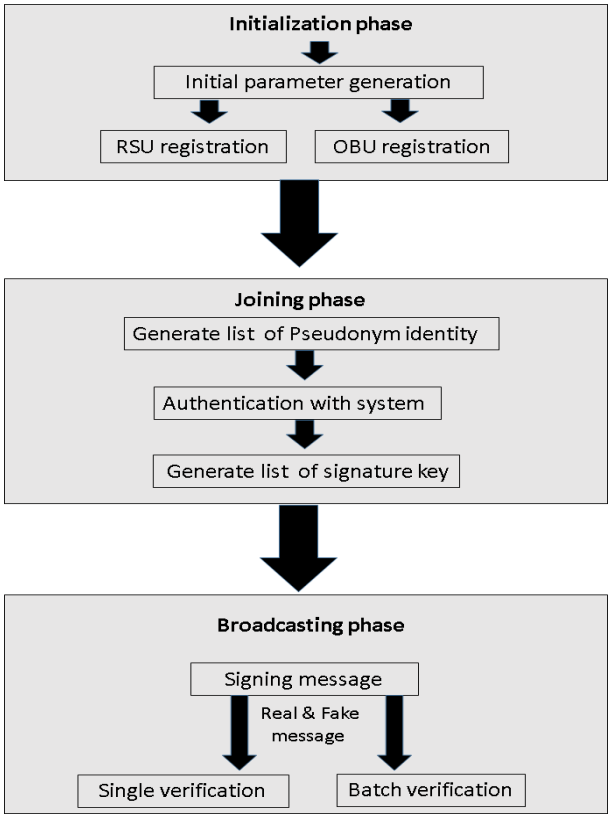
A security analysis of the proposed scheme is provided in this section to clarify that our scheme is secure under a Burrows– Abadi–Needham (BAN) logic, random oracle model and proof of security. We also provide the requirements for secu-rity and privacy in this paper.

3. SYSTEM ARCHITECTURE AND PRELIMINARIES

In the following parts, the necessary mathematical tools used in this study are introduced. Then, the model for vehicu- lar communication and the adversary models are discussed. Finally, the security and privacy requirements for the pro- posed scheme are described. Table 1 contains some notationand their description.

Notation	Descriptions
E	An elliptic curve
G	An additive group based on E
a, b	Two large prime number
p	large prime number
P	The base generator P∈ G
h_1, h_2, h_3	Three one-way hash function
RID_{RSU_j}, RID_i	Real identity of the RSU and vehicle
PW_i	Password of driver
$x_{pri}^{TA}, s_{Pri}^{dom_i}$	The private master key of the system and $domain_i$
$P_{Pub}^{TA}, P_{Pub}^{dom_i}$	The public key of the TA and $domain_i$
r, z	Random integer
$\|$	Concatenation operation
$\oplus$	XOR operator
L_{PID_i}	List of OBU_i's local Pseudo identities
L_{SK_i}	List of OBU_i's local Private keys
R_1, L_1	Share secret key

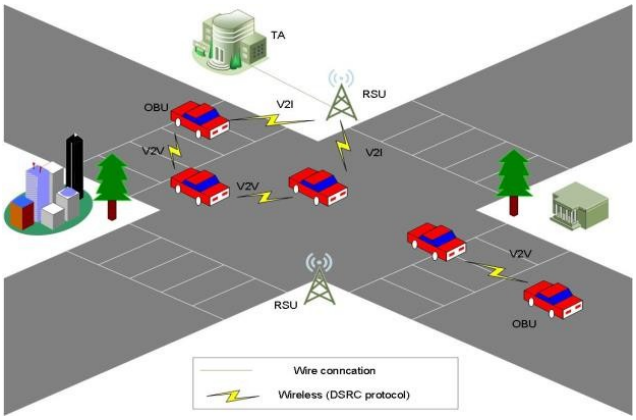
Miller [26] suggested ECC, an algorithm that is widely used to provide asymmetrical encryption in an elliptical curve. This algorithm has smaller key lengths than the same security level as other encryption algorithms.



Definition 1 (Elliptic Curve): Let F_p be a finite field, and a large prime number p is the order of F_p. E is a parameter, the vehicle must establish a shared authentication with the nearest RSU in any domain to begin transmission and validate operations. Then, the TA will confirm the authenticity of the OBU via the private key of the system. Thereafter, the RSU generatesa list of signatures that can be used in the selected timestamp, and then sends them securely to the OBU. n is a level of security anonymity, that is, the number of pseudo identities that a vehicle may unrepeatable in a region enclosedby the RSU [27]. Finally, the OBU uses the signature list until the time list expires. Figure 3 briefly describes the proposed scheme phases. The following subsection explainsthree phases in detail.

INITIALIZATION PHASE

During this phase, the TA creates system parameters to use the following steps:



elliptic curvedefined as: $y^2 = + x^3 +$ ax b mod ∈ p.

a, b F_p are constants. A group G_q is defined on E, whose order is q and generator is P. The set contains an infinity point O.

Scalar multiplication. Let P ∈ Gq, n ∈ Z ∗, such that the scalar multiplication is x . P = P + P+P +P (x times).

Definition 2: Elliptic curve discrete logarithm problem (ECDLP): is computationally infeasible. E has two random.

4. PROPOSED SCHEME

The proposed scheme has three phases: initialization, joining, and broadcasting. In this scheme, after TA generates the initial public parameters of the system, the TA calculates the private and public keys for the *domain*$_i$, which contains several registered RSUs from the registration list located nearby in a specific area (e.g., city). The TA also stores the registered OBUs to the vehicle registration list.

In the second phase, after the OBU produces n pseudo ID list with its real identity and public TA

5. SYSTEM REQUIREMENTS

HARDWARE SYSTEM CONFIGURATION

- processor : Pentium – IV
- RAM : 4 GB
- Hard Disk : 20 GB

SOFTWARE SYSTEM CONFIGURATION

- Operating System : Windows 7 or 8
- Software : python Idle

MODEL DESIGN

- Data set collection
- Training of images
- Traffic flow rate analysis
- Traffic congestion

6. CONCLUSION AND FUTURE WORK

Intelligent Transport System (ITS) has been gaining momentum as more elements in a transport system are becoming more connected. In line with this, VANETs are becoming popular and greatly contribute to ITS. The specifications for contents and contextual privacy must be met to protect privacy vehicles in terms of identity and location as suscepti- ble information. In this paper, we have proposed a schemeto ensure these requirements are met. The scheme ensures privacy of data through signing and verifying traffic-relatedmessages, which are protected by the proposed VPPCS scheme. It also meets the requirement of all contextual privacy on the grounds of the injection for fake traffic- related messages. Security and performance analyses were performed to validate the proposed scheme. The security analysis shows that VPPCS can withstand model security attacks and satisfy all privacy requirements. The performance evaluation reveals that the scheme proposed by VPPCS is VANET compatible and that our VANET scheme is efficient in terms of computational cost and communication over- head. The balance between privacy and performance was also emphasized.

When the pseudonym set is expired, the vehicle removes the old set and then requests to obtain a new set. Conse- quently, there is no accumulated storage, which leads to the overhead increased. In future research, the main focus of thenext paper is to address the overhead of storage in the VANET system. Besides, we will carry out simulation experiment through simulation platform such as OMNET and SUMO to demonstrate the performance of the work.

7. REFERENCES

[1] V. Hoa La and A. Cavalli, ''Security attacks and solutions in vehicular ad hoc networks: A survey,'' *Int. J. AdHoc Netw. Syst.*, vol. 4, no. 2, pp. 1–20, Apr.2014.

[2] F. Qu, Z. Wu, F. Wang, and W. Cho, ''A security and privacy review of VANETs,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 2985–2996, Dec. 2015.

[3] M. Al Shareeda, A. Khalil, and W. Fahs, ''Realistic heterogeneous genetic- based RSU placement solution for V2I networks,'' *Int. Arab J. Inf. Tech- nol.*, vol. 16, no. 3, pp. 540–547, 2019.

[4] M. Raya and J.-P. Hubaux, ''Securing vehicular ad hoc networks,'' *J. Com-put. Secur.*, vol. 15, no. 1, pp. 39–68, 2007.

[5] M. Al Shareeda, A. Khalil, and W. Fahs, ''Towards the optimization of road side unit placement using genetic algorithm,'' in *Proc. Int. Arab Conf. Inf. Technol. (ACIT)*, Nov. 2018, pp. 1–5.

[6] M. R. Jabbarpour, H. Zarrabi, R. H. Khokhar, S. Shamshirband, and K.-K.-R. Choo, ''Applications of computational intelligence in vehicle traffic congestion problem: A survey,'' *Soft Comput.*, vol. 22, no. 7,pp. 2299–2320, Apr. 2018.

[7] Z. Zhou, C. Gao, C. Xu, Y. Zhang, S. Mumtaz, and J. Rodriguez, ''Social big-data-based content dissemination in Internet of vehicles,'' *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 768–777, Feb. 2018.

[8] S. Zeadally, R. Hunt, Y.-S. Chen, A. Irwin, and A. Hassan, ''Vehicular ad hoc networks (VANETs): Status, results, and challenges,'' *Telecommun. Syst.*, vol. 50, no. 4, pp. 217–241, 2012.

[9] D. Jacobs, K.-K.-R. Choo, M.-T. Kechadi, and N.-A. Le-Khac, ''Volk- swagen car entertainment system forensics,'' in *Proc. IEEE Trust- com/BigDataSE/ICESS*, Aug. 2017, pp. 699–705.

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Realtime Facial Emotion Recognition System

AFSAL SUBAN[1], HARIPRASATH MURUGAN[2], KIRUTHICKROSAN ANBU KUMAR[3], KRISHNAMOORTHY MUTHU[4]
[1]Assistant Professor, [2,3,4]Student
[1,2,3,4]Department of CSE, AVS Engineering College, Salem, Tamil Nadu, India

ABSTRACT

Face detection has been around for ages. The main objective of face recognition is to authenticate and identify the countenance. Takinga step forward, human emotion detection is the need of the hour so that modern artificial intelligent systems can get reactions fromface. Facial emotion detection can be used to understandthe human behavior, detection of mental disorders and synthetic human expressions. This can be doneby using machine learning algorithms used in facial recognition for accurate identification and detection. However, the countenance are captured in real time and processed using haar cascade detection. The project work is defined in three different phases where within the first phase, Face is detected from the camera and within the second phase, the captured input is analysed using the features with support of keras convolutional neuralnetwork model. Within the last phase, Face is authenticated to classify the emotions of human as happy, neutral, angry, sad, disgust, fear and surprise. The proposed work presented is simplified in three phases such as detection of face, recognition and classification of emotion. In this project Open CV library, Keras, Tensorflow, Pandas, Numpy, Dataset, Jupyter Notebook and Python Programming is used.

Index Terms— Realtime System.

1. INTRODUCTION

Recognizing facial emotions has become a major issue in many applications today. The research on facial emotion recognition has gained a lot of momentum over the past few years. The state of human emotions is identified using facial emotion recognition (e.g. neutral, happy, sad, surprise, fear, anger, disgust, contempt) based on the stream of images fed in by a video.

There has been growing interest in making machines act as close as possible to actual human beings. To make their actions replicate those of humans and add a touch of human feelings in each of these actions. It has been argued that for there to be a proper human-computer interaction, the computer has to interact in a natural way, similar to that whentwo humans interact. Other fields where emotion recognition can prove to be useful are online teaching, product marketing, health industry, and several others to determining whether certain things are liked or not and what is the reaction of different people towards different stimuli. Humans express their thoughts and feelings in multiple ways. The most important being speech, followed up by their display of emotions. Emotions typically manifest themselves via multiple means, be it touch, visual or physiological. The mostappropriate method of detecting human emotions would be todetermine their emotion from the visual cues displayed by a human. It is widely accepted that the slightest of change in emotions become visible on the faces of humans and hence developing a system which detects the emotion from an individual's face could prove to be an invaluable tool.

The work done by us in this project enables proposes a system which allows for real-time emotion detection by using a video stream input from the user's webcam. We initially focus on detecting the face from the video using the Haar cascade classifiers. We then make use of convolutional neural networks for determining emotions. The dataset used by us in this project is the FER2013 dataset. The dataset originally was in the form of raw pixels but we had to convert them into actual images and use those images for training our model.

This greatly helped in improving the performance of our model. The software used by us in this project is OpenCV and TensorFlow. OpenCV is used for operating the webcam and for face detection, while TensorFlow was used for training the CNN for emotion detection.

2. INPUT AND OUTPUT DESIGN

Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easierand to be free from errors. The data entry screen is designed in such a way that all the data manipulatescan be performed. It also provides record viewing facilities.

3. IMPLEMENTATION & TESTING

TEST APPROACH: is the test strategy implementation of a project, defines how testingwould be carried out. Test approach has two techniques:

- ***Proactive*** **-** An approach in which the test design process is initiated as early as possible in order to find and fix the defects before the build is created.

- ***Reactive -*** An approach in which the testing is not started until after design and coding

UNIT TESTING: Unit testing, a testing technique using which individual modules are tested to determine if there are any issues by the developer himself. It is concerned with functional correctness of the standalone modules. The main aim is to isolate each unit of the system to identify, analyze and fix the defects.

INTEGRATED TESTING: Integration testing is the second level of the software testing process comes after unit testing. In this testing, units or individual components of the software are tested in a group. The focus of the integration testing level isto expose defects at the time of interaction between integrated components or units. Unit testing uses modules for testing purpose, and these modules are combined and tested in integration testing. The Software is developed with a number of software modules that are coded by different coders or programmers. The goal of integration testing is to check the correctness of communication among all the modules. Once all the components or modules are working independently, then we need to check the data flow between the dependent modules is known as integration testing

BETA TESTING: Beta testing also known as user testing takes place at the end users site by the end users to validate the usability, functionality, compatibility, and reliability testing. Beta testing adds value to the software development life cycle as it allows the "real" customer an opportunity to provide inputs into the design, functionality, and usability of a product. These inputs are not only critical to the success of the product but also an investment into future products when the gathered data is managed effectively.

4. SYSTEM ANALYSIS

A. PROPOSED SYSTEM - Systems design is aprocess that defines architecture, components, modules, interfaces, and data requirements. System design can be viewed as a system theory application for product development. The face detection technology that helps locate human face in digital images and video frames. The object detection technology that deals with detecting instances of objects in digital image and videos. The proposed automated recognition system can bedivided into five main modules:

Image Capture A camera is placed away from the entrance to capture an image of the front of the student. And a further process goes for face detection.

Face Detection and Facial Features The appropriate and effective facial detection algorithm constantly improves facial recognition. Several facial algorithms such as face-to-face geometry, construction methods, Face geometry-based methods, Feature Invariant methods, Machine learning based methods. Out of all these methods Viola and Jones proposed a framework that gives a high detection rate and is also fast. Viola-Jones detection algorithm is fast and robust. So we chose Viola-Jones face detection algorithm, which uses Integral Image and AdaBoost learning algorithm as classier. We have observed that this algorithm yields better results in a variety of lighting conditions.

Pre-Processing Extracting the face features it is called pre-processing. This pre-processes step involves specifying the extracted facial image and transforms to 100x100. Histogram Equalization is the most commonly used Histogram Normalization technique. This improves the contrast of the image as it extends beyond the intensity of the image, making it even more clear and constraint.

Database Development As we choose biometric based system every individual is required. This database development phase consists of an image capture of each individual and extracting the biometric feature, and then it is enhanced using preprocessing techniques and stored in the database.

Post-Processing In the proposed system, after recognizing the faces of the person, the names are show into a video output. The result is generated by exporting mechanism present in the database system. These generated records can be seen in real time video. This ensures that person whose faces are not recognized correctly by the system have to check in database. And thus, giving them the ability to correct the system and make it more stable and accurate.

5. SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS

- Processor : 2GHz Intel Core i3
- Hard Disk : 2 GB or more
- Memory (RAM) : 4 GB
- Webcam

SOFTWARE REQUIREMENTS

- Programming Language : Python
- Front End : HTML/CSS
- Back End : Flask
- Operating System: Microsoft Windows 7/8/10(32- or 64-bit)
- Database : MySQL
- Software : Anaconda

6. DESIGN

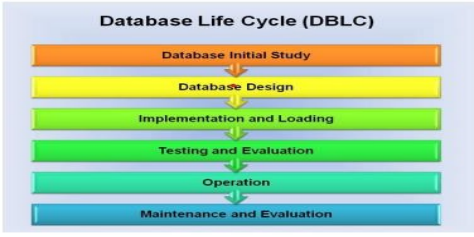
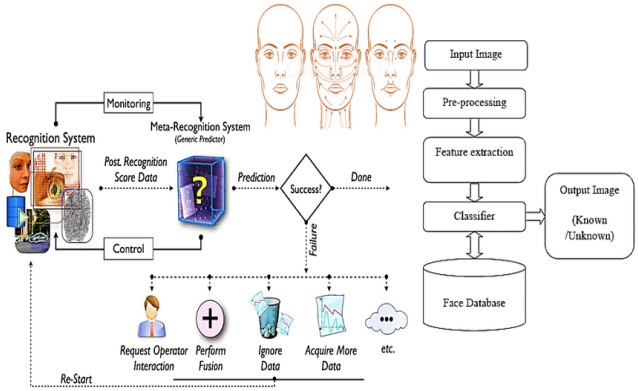


Figure- Database Life Cycle

A.SYSTEM DESIGN
B.SYSTEM ARCHITECTURE DIAGRAM



7. CONCLUSION

Nowadays I had trained my model on 35887 images in which 28821 images are belonging to train folder and 7066 images belonging to validation folder. I have successfully implemented the Realtime Facial Emotion Recognition System with the Training accuracy of 96.78% and validation accuracy of 67.03%. I have implemented all above test cases successfully. I achieved clean interfaces that further enhance the user experience. User can identify their emotions efficiently. Some Activities according to emotion will be suggested to user so that user can do that activity and feel good. The activity link is provided. So we can conclude that by using this application user can find their emotions and make themselves free from stress, tension and depression. For example if user is sad than they canwrite their feelings into diary, if user is angry then they can do some meditations, if user is happy than they can perform some dance move so on. In this way user can spend some time for themselves.

8. REFERENCES

[1] Vandal, T., McDuff, D., & El Kaliouby, R. Event detection: Ultra large-scale clustering of facial expressions. In Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on Vol. 1, pages 1–8,2015.

[2] Irene Kotsia, Ioan Buciu, and Ioannis Pitas. An analysis of facial expression recognition under partial facial image occlusion. Image and Vision Computing, 26(7):1052–1067, 2008.

[3] Nakasone, Arturo, Helmut Prendinger, and Mitsuru Ishizuka. "Emotion recognition from electromyography and skin conductance." Proc. of the 5th international workshop on biosignal interpretation. 2005.

[4] Mehendale, N. Facial emotion recognition using convolutional neural networks (FERC). SNAppl. Sci. 2, 446 (2020). https://doi.org/10.1007/s42452-020-2234-1

[5] Saravanan, Akash & Perichetla, Gurudutt & K.s, Gayathri. (2019). Facial Emotion Recognition using Convolutional Neural Networks.

[6] G. A. R. Kumar, R. K. Kumar and G. Sanyal, "Facial emotion analysis using deep convolution neural network," 2017 International Conference on Signal Processing and Communication (ICSPC), 2017, pp. 369-374, doi: 10.1109/CSPC.2017.8305872.

[7] Dachapally, Prudhvi. (2017). Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units.

[8] Helaly, Rabie & Hajjaji, Mohamed & Faouzi, M'Sahli & Abdellatif, Mtibaa. (2020). Deep Convolution Neural Network Implementation for Emotion Recognition System. 261-265. 10.1109/STA50679.2020.9329302.

[9] Ghaffar, Faisal. (2020). Facial Emotions Recognition using Convolutional Neural Net.

[10] Li, Chieh-En James and Zhao, Lanqing, "Emotion Recognition using Convolutional Neural Networks" (2019). Purdue Undergraduate Research Conference. 63. https://docs.lib.purdue.edu/purc/2019/Posters/63

[11] M. A. Ozdemir, B. Elagoz, A. Alaybeyoglu, R. Sadighzadeh and A. Akan, "Real Time Emotion Recognition from Facial Expressions Using CNN Architecture," 2019 Medical Technologies Congress (TIPTEKNO), 2019, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895215.

[12] Tanner Gilligan, Baris Akis Emotion AI, Real- Time Emotion Detection using CNN " http://web.stanford.edu/class/cs231a/ prev_projects_2016/emotion-ai-real.pdf

[13] R. Pathar, A. Adivarekar, A. Mishra and A. Deshmukh, "Human Emotion Recognition using Convolutional Neural Network in Real Time," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 2019, pp. 1-7, doi: 10.1109/ICIICT1.2019.8741491.

[14] Huang, Y.; Chen, F.; Lv, S.; Wang, X. Facial Expression Recognition: A Survey. Symmetry 2019, 11, 1189 https://doi.org/10.3390/sym11101189

❑❑❑

Recent Developments in Detection of Central Serous Retinopathy Through Imaging and Artificial IntelligenceTechniques—A Review

KARTHIKRAJA. M[1], PRATHAP. M[2], SRI VIJAYARAGAVI. P[3], YOGALAKSHMI. R[4]
[1]Assistant Professor, [2,3,4]Student
[1,2,3,4]Department of CSE, Ganesh College of Engineering, Salem, Tamil Nadu, India

ABSTRACT

Central Serous Retinopathy (CSR) or Central Serous Chorioretinopathy (CSC) is a significant disease that causes blindness and vision loss among millions of people worldwide. It transpires as a result of accumulation of watery fluids behind the retina. Therefore, detection of CSR at early stages allows preventive measures to avert any impairment to the human eye. Traditionally, several manual methods for detecting CSR have been developed in the past; however, they have shown to be imprecise and unreliable. Consequently, Artificial Intelligence (AI) services in the medical field, including automated CSR detection, are now possible to detect and cure this disease. This review assessed a variety of innovative technologies and researches that contribute to the automatic detection of CSR. In this review, various CSR disease detection techniques, broadly classified into two categories: a) CSR detection based on classical imaging technologies, and b) CSR detection based on Machine/Deep Learning methods, have been reviewed after an elaborated evaluation of 29 different relevant articles.

Index Terms— Central Serous Retinopathy, Deep Learning, Fundus Images, MachineLearning, Optical Coherence Tomography Images.

1. INTRODUCTION

The retina is located behind the eyeball near the optic nerve and comprises a thin layer of tissue . It obtains the focused light from the eye-lens, converts it into neural signals, and imparts signs to the brain for visual recognition. The retina processes light using a layer of photoreceptor cells. These are light-sensitive cells responsible for detecting visual characteristics, such as color and light intensity.

Subsequently, the data accumulated by the photoreceptor cells are sent to the brain through the optic nerve for optical recognition. Therefore, the retina plays a crucial role in image processing for the human brain recognizes and distinguishes various surrounding objects and names them. Since any damage to the retina may have severe ramifications to our ocular abilities.

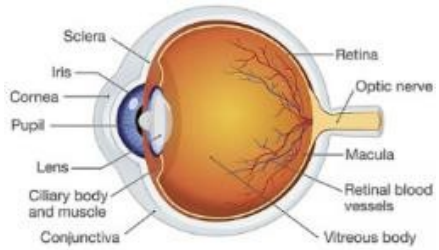


FIGURE 1: Anatomy of the Human Eye, depicting various parts includingthe Retina

methods in a specific retinal disorderknown as the Central Serous Retinopathy (CSR).The The interpretation of CSR in classical terms is referred to as acute CSR. A patient with intense CSR may encounter obscured vision, decrease in contrast sensitivity and shading vision, metamorphopsia, and minor hyperopic move. Traditionally there is a central serous separation of focal retina, sometimes with dull yellow stores and in a few cases with serous RPE separation. In several cases of CSR, a permanent sub-retinal fluid accumulates for three months or more: resulting in permanent visual symptoms. Such cases of CSR frequently experience a fluctuating degree of sub- retinal fluid. However, most cases recover automatically, whereas some patients may experience chronic CSR.

Nevertheless, their visual acuity remains typically steady. The chronicity of these patients depends upon the time duration of CSR, and it usually takes 3 to 6 months in case of acute CSR. Contrarily, in chronic CSR patients, the symptoms of morphological changes and an increasedrisk of CNV have been observed. Patients with CSC are generally of the age of 25 and 50, in which men are afflicted far more frequently than women. These patients normally have symptoms such as the grumblings of unexpected beginning, contortion, and focal vision blurring.

Visual acuity ranges between 6/5 and 6/60, but its usual range is 6/9 to 6/12 [34]. In the case of CSR, it is generally a self- constraining disease with unconstrained resolution having boundaries of 3–4 months. Historically data reveals that nearly half of the CSR.

Patients may experience recurrences of thedisease within a year, causing the patient to undergo various treatment procedures, which may last for three months in chronic CSR, recurrent CSR, and first- time CSR patients. Some standard CSR treatments include the Micro Pulse Laser Treatment (MPLT), the Transpupillary Thermo treatment

(TTT), the Photodynamic therapy (PDT), and the Intravitreal anti-Vascular Endothelial Growth Factor (anti-VEGF). These treatment methods are based on the following points:

- A majority of the population automatically recuperates within 4 to 6 months without requiring any specificmedicaments.
- If a patient's CSR lasts for a year, then some treatment may be required.
- In rare cases, if CSR lasts for more than a year, an ophthalmologist may opt for specific treatments such as RPE detachment or bullous retinal detachment.
- There have been approximately 40 articles published related to deep learning and medical imaging. Among these 40 articles, only 29 articles are associated with Central Serous Retinopathy (CSR). These articles, along with their methodologies, advantages, and limitations, are discussed in this review. As mentioned above, CSR detection is performed through various imaging technologies.

2. IMAGING TECHNOLOGIES FOR DETECTION OF CSR

The traditional imaging technologies for CSR detection include Fundus Photography, Fluorescein Angiography (FA), and Optical Coherence Tomography (OCT). These technologies are discussedin the following subsections.

THE FUNDUS PHOTOGRAPHY

Fundus Photography is a process of obtaining the retinal red free image and is considered an alternative to OCT imaging. This technique is based on the statistical approach and requires advancement in contrast to its forerunner, color photography film. Similarly, the digital image of retina provides quick towering- resolution and consistent image, and it is accessible instantly and manageable for the development of an image.

Moreover, Fundus photography is regularly employed for ailment records and clinical examination, along with potential usage for tolerant training and telehealth. Additionally, the images generated through Fundus techniques can incorporate average and extensive views [3]. Figures 2 and 3 depict retinal scans obtained via fundus photography [36], [37]. In Figure 2, the normal indications of a healthy eye have been depicted, whereas, in Figure 3, the dark spot of a blister of fluid caused by CSR disease has been shown.

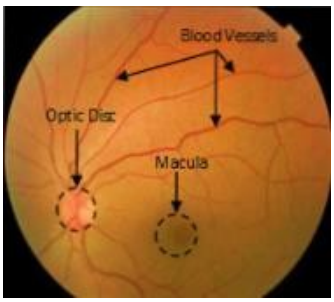


Figure 2: Normal Retinal Fundus Scan

FLUORESCEIN ANGIOGRAPHY (FA)

Fluorescein Angiography (FA) is an effective imaging technique in which fluorescent dye is injected into blood vessels of patient's eyes in order to capture their clear images. The main objective ofthis technique is to highlight the blood vessels to form a clear and visible image. The patient is normally prescribed with primary care before initiating the FA procedure to ensure a satisfactory blood stream in the veins. In addition, the physicians propose primary care to analyze further the eye issues, including macular degeneration, diabetic retinopathy, or Central Serous Retinopathy (CSR).

Optical Coherence Tomography (OCT) appeared as an advanced automation technology used for detection and diagnosis of different diseases developed in the early 1990s. Above mentionedapproach is similar to ultrasound imaging. However, instead of making use of sound it uses the light. The combination of catheters and endoscopes with OCT produces high-resolution imaging of the organ system. OCT can provide tissue images intangible and situ form. segment, Interface between IS and OS, Photoreceptor inner segment and Outer Photoreceptors [38]. These layers are present in retina OCT as shown in Figure6.

OCT IMAGING DATASETS

In this study, two publicly available datasets associated with CSR are analyzed. A typical dataset contains a range of records, and these CSR datasets consist of the repositories of OCT and fundus images. Several researchers normally access and utilize these publicly accessible datasets, which can be easily retrieved using their specific links. In various experimental studies, a portion of the all- out images is normally downloaded from these datasets in order totrain new Machine Learning models and algorithms and achieve testing goals. In the following list, we briefly describe two publicly available datasets along with their accessibility.

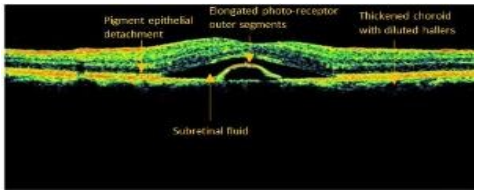


FIGURE 7. Retinal OCT image with CSR disease

OCTID DATASET

The OCT imaging database is an open- source OCT imaging database, which is accessible at the website of University ofWaterloo, Canada.

3. LITERATURE REVIEW

In this section, a comprehensive literature review has been compiled based on thefollowing steps:

- Identification of relevant and advanced research articles.

- American journal of ophthalmology, (https://www.ajo.com/)
- Articles emphasizing on thebasic
- A comprehensive search strategy that identifies the research problem.
- Extraction of the desired data from the selected researcharticles.
- Validation and fact-checking ofthe collected data.
- Representation of data for better visualization and The literature review articles have been searched from the following reputable and well- recognized scientific publishersand organizations:

4. DISCUSSION

DL technique is quite beneficial for fundus image sorting procedures, and it yields better results accuracy results than the classical ML approach. How- ever, the results largely depend on the size ofdatasets and the complexity of theunderlying algorithms.

Figure 13 shows total no of images used inthe literature with respect to each classifier.

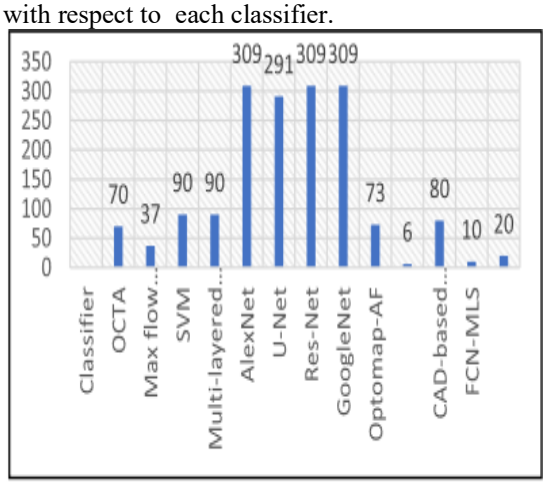


FIGURE 13. Comparison of literature based on validation data images.

5. CONCLUSION

In this study, a detailed review of automatic detection of CSR using AI techniques has been presented. These automatic detection methods assist in detecting the disease of vision loss and blindness as a result of CSR. These methods were categorized into classical imaging and ML/DL based techniques. The significance of each ML and DL methodology was ascertained to perform a better analysis of CSR for research community and ophthalmologists. The algorithm introduced by Hassan et al. [84] is the most promising algorithm as it gives99.8% accuracy. This open-source CSR detection model will provide agility, sustainability flexibility and cost effectiveness, thus it can serve the mankind in a much better way. Detectionof CSR is evolving, using Machine Learning/Deep Learning Algorithms and imaging technologies.

6. REFERENCES

[1] C. E. Willoughby, D. Ponzin, S. Ferrari, A. Lobo, K. Landau, and Y. Omidi, ''Anatomy and physiology of the human eye: Effects of mucopolysac- charidoses disease on structure and function—Areview,''Clin. Experim.Ophthalmol., vol. 38, pp. 2–11, Aug. 2010, doi: 10.1111/j.1442- 9071.2010.02363.x.

[2] A. A. Dahl and T. R. Gest, ''Retina anatomy,'' Medscape Reference, pp. 1–4, 2015. Accessed: Oct. 10, 2020. [Online]. Available: https://www.nvisioncenters.com/education/eye-structure-anatomy/

[3] M. U. Akram, S. Akbar, T. Hassan, S. G. Khawaja, U. Yasin, and I. Basit, ''Data on fundus images for vessels segmentation, detection of hypertensive retinopathy, diabetic retinopathy and papilledema,'' Data Brief, vol. 29, Apr. 2020, Art. no. 105282, doi: 10.1016/j.dib.2020.105282.S. Akbar, T. Hassan, M. U. Akram, U.

[4] U. Yasin, and I. Basit, ''AVRDB: Annotated dataset for vesselsegmentation and calculation ofarteriovenous ratio,'' in Proc. Int. Conf.Image Process., Comput. Vis., Pattern Recognit. (IPCV), 2017, pp. 129–134.

[5] S. Akbar, M. Sharif, M. U. Akram, T. Saba, T. Mahmood, and M. Kolivand, ''Automated techniques for blood vessels segmentation through fundus retinal images: A review,'' Microsc. Res. Technique, vol. 82, no. 2, pp. 153–17Feb.

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Investigations on Personalized Recommendation System Based on Collaborative Filtering for IOT Scenarios

K. LAKSHMANAN[1], VENNILA.V[2]
[1]Assistant Professor, [2]Student
[1,2]Department of CSE, Annai Mathammal Sheela Engineering College, Erumapatty, Tamil Nadu, India

ABSTRACT

Social voting is an emerging new feature in online social networks. It poses unique challenges and opportunities for recommendation. In this paper, we develop a set of matrix factorization (MF) and nearest-neighbor (NN)-based recommender systems (RSs) that explore user social network and group affiliation information for social voting recommendation. Through experiments with real social voting traces, we demonstrate that social network and group affiliation information can significantly improve the accuracy of popularity-based voting recommendation, and social network information dominates group affiliation information in NN based approaches. We also observe that social and group information is much more valuable to cold users than to heavy users. In our experiments, simple metapath based NN models outperform computation-intensive MF models in hot-voting recommendation, while users' interest for nonhot votings can be better mined by MF models. We further propose a hybrid RS, bagging different single approaches to achieve the best top-k hit rate.

Index Terms.***

1. INTRODUCTION

ONLINE social networks (OSN), such as Facebook and Twitter, facilitate easy information sharing among friends. A user not only can share her updates, in forms of text, picture, and video, with her direct friends, but also can quickly disseminate those updates to a much larger audience of indirect friends, leveraging on the rich connectivity and global reach of popular OSNs. Many OSNs now offer the social voting function, through which a user can share with friends her opinions, e.g., like or dislike, on various subjects, ranging from user statuses, profile pictures, to games played, products purchased, websites visited, and so on. Taking like dislike type of votings one step further, some OSNs, e.g., Sina Weibo, empower users to initiatetheir own voting campaigns, on any topic of their interests, with usercustomized voting options. The friends of a voting initiator can participate in the campaign or retweet the campaign to their friends.

Also, the purpose of initializing a voting is to engage people to express their opinions. Thus, the topics covered in online social votings are generally more engaging than other applications in OSNs. Section III presents some interesting statistics of our online social voting data trace.

2. LITERATURE SURVEY

"Secure Friend Discovery in Mobile Social Networks," AUTHORS: W. Dong, V. Dave, L. Qiu, and Y. Zhang,
Mobile social networks extend social networks in the cyberspace into the real world by allowing mobile users to discover and interact with existing and potential friends who happen to be in their physical vicinity. Despite their promise to enable many exciting applications, serious security and privacy concerns have hindered wide adoption of these networks. To address these concerns, in this paper we develop novel techniques and protocols to compute social proximity between two users to discover pote contributions. First, we identify a range of potential attacks against friend discovery by analyzing real traces. Second, we develop a novel solution for secure proximity estimation, which allows users to identify potential friends by computing social proximity in a privacy-preserving manner ntial friends, which is an essential task for mobile social networks. We make three major A distinctive feature of our solution is that it providesboth privacy and verifiability, which are frequently at odds in secure multi party computation. Third, we demonstrate the feasibility and effectiveness of our approaches using real implementation on smart phones and show it is efficient in terms of both computation time and power consumption.

"Seer: A Secure and Efficient Service Review System for Service-Oriented Mobile Social Networks"
AUTHORS: X. Liang, X. Li, R. Lu, X. Lin, and X. Shen
In this paper, we consider service-oriented mobile social networks (S-MSNs) and propose a Secure and Efficient service Review (SEER) system to enable user feedback. Each service provider independently maintains a SEER system for itself, which collects and stores user reviews about its services without requiring any central trusted authority. The service reviewscan then be made available to interested users in making wise service selection decisions. We identify three unique service review attacks and then develop sophisticated security mechanisms for SEER to deal with these attacks. Specifically, SEER enables users to distributed and cooperatively submit their reviews in an integrated chain form by using hierarchical and aggregate signature techniques. It discourages service providers to reject, modify or delete their reviews.

3. SYSTEM ANALYSIS

EXISTING SYSTEM:

Online social voting has not been much investigated to our knowledge. Social voting as a newsocial network application has not been studied much in the existing literature. Compared withtraditional items for recommendation, the uniqueness of online social voting lays in its socialpropagation along social links. Also, the purpose of initializing a voting is to engage people toexpress their opinions. Thus, the topics covered in online social votings are generally moreengaging than other applications in OSNs

DISADVANTAGES:

- Online social voting has not been much investigated
- Doesn't satisfy the online social user¡¦s requirement.

PROPOSED SYSTEM:

To develop MF-based and NN-based RS models to learn user-voting interests by simultaneouslymining information on user-voting participation, user friendship, and user group affliction. We show through experiments with real social voting traces that both social network informationand group affiliation information can be mined to significantly improve the accuracy ofpopularity-based voting recommendation. We show that simple met path-based NN modelsoutperform computation-intensive MF models in hot-voting recommendation, while users' interest for no hotvoting's can be better mined by MF models.

ADVANTAGES:

- Users can be easily overwhelmed by various voting's that were initiated, participated
- Recommender systems (RSs) deal with information overload by suggesting to users theitems that are potentially of their interests.

IMPLEMENTATION MODULES:

- User Module
- Trusted Authority Module
- Sybil Attack Detection
- Token Synchronization
- Voting on Recommendation

4. MODULE DESCRIPTION

USER MODULE:

The user can register and he/she can login. After login that user generates multiple reviews toward a vendor in a predefined time slot with different Pseudonyms and users are ableto frequently change their pseudonyms to prevent the linkage of their behaviors at different time/location.

TRUSTED AUTHORITY MODULE:

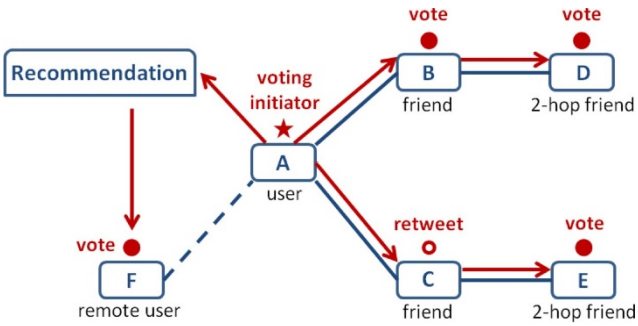
Trusted Authority, to receive user feedback, known as service reviews or simply reviews, such as compliments and complaints about their services or products. By using the TSE, the service providers learn the service experiences of the users are able to improve their service strategy in time. In addition, the collected reviews can be made available to the public, which enhances service advertising and assists the users in making wise service selections. The TSE is often maintained by a third trusted authority that is trusted to host authentic reviews.

SYBIL ATTACK DETECTION:

The further extend the TSE to a Sybil-resisted TSE, named TSE, which effectivelyprevents the Sybil attacks. We define two types of Sybil attacks: The Sybil attack 1 is launchedby a group of registered users. They aim at telling other users the bad service from a vendor while the service of the vendor is good. With the valid registration, these malicious users are ableto leave false reviews toward a specific vendor.

TOKEN SYNCHRONIZATION:

The chain structure requires reviews to be submitted sequentially. The tokens will then be circulated among users according to their local decision on token forwarding. A user cannot submit a review unless it currently holds one of the tokens. A token may be lost due to user mobility or malicious dropping.



SYSTEM ARCHITECTURE

SYSTEM DESIGN

5. SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 500 GB.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- RAM : 4 GB.

SOFTWARE REQUIREMENTS:

- Operating system : Windows -7/8/10.
- Coding Language : Java 1.8
- IDE Tools : NetBeans 8.0
- Database : MYSQL 10.0

6. CONCLUSION

The Proposed TSE system for S-MSNs. The system engages hierarchical signature and aggregate signature techniques to transform independent reviews into structured review chains. This transformation involves distributed user cooperation, which improves review

integrity and significantly reduces vendors' modification capability. The presented three review attacks and shown that the TSE can effectively resist the review attacks without relying on a third trusted authority. Considered the notorious Sybil attacks and demonstrated that such attacks cause huge damage to the TSE.

7. REFERENCES

[1] W. Dong, V. Dave, L. Qiu, and Y. Zhang, "Secure Friend Discovery in Mobile Social Networks," Proc. IEEE INFOCOM, pp. 1647-1655, 2011.

[2] X. Liang, X. Li, R. Lu, X. Lin, and X. Shen, "Seer: A Secure and Efficient Service Review System for Service-Oriented Mobile Social Networks," Proc. IEEE 32nd Int'l Conf. Distributed Computing Systems (ICDCS), pp. 647-656, 2012.

[3] X. Liang, X. Li, T. Luan, R. Lu, X. Lin, and X. Shen, "Morality- Driven Data Forwarding with Privacy Preservation in Mobile Social Networks," IEEE Trans. Vehicular Technology, vol. 61, no. 7, pp. 3209-3222, Sept. 2012.

[4] T.H. Luan, L.X. Cai, J. Chen, X. Shen, and F. Bai, "VTube: Towards the Media Rich City Life with Autonomous Vehicular Content Distribution," Proc. IEEE CS Eighth Ann. Conf. Sensor, Mesh and Ad Hoc Comm. Networks (SECON), pp. 359-367, 2011.

❑❑❑

Efficient Traffic Signs Recognition Based on CNN Model for Self-Driving Cars

M.SARANYA[1], P.AKILA[2], M.MENAKA[3], V.VIJAYAKUMAR[4]
[1]Assistant Professor, [2,3,4]Student
[1,2,3,4]Department of CSE, Annapoorana Engineering College, Salem, Tamil Nadu, India

ABSTRACT

Due to the unavailability of Vehicle-to-Infrastructure (V2I) communication in current transportation systems, Traffic Light Detection (TLD) is still considered an important module in autonomous vehicles and Driver Assistance Systems (DAS). To overcome low flexibility and accuracy of vision-based heuristic algorithms and high-power consumption of deep learning-based methods, we propose a lightweight and real-time traffic light detector forthe autonomous vehicle platform. Our model consists of a heuristic candidateregion selection module to identify all possible traffic lights, and a lightweight Convolution Neural Network (CNN) classifier to classifythe results obtained. The main goal of this research is to detect traffic light in real-time for autonomous vehicles. Apart from taking decisions to navigate in the right manner the autonomous vehicles important task is to detect traffic lights, so that it can obey the traffic rules with sufficient precision. The work carried out in this research makes use of two Artificial Intelligence technique, these techniques are compared in accomplishing the task of traffic light detection in real time.

Index Terms - Machine Learning, Deep Learning, Traffic Signs Recognition, Convolu-Tional Neural Networks, Autonomous Driving, Self-DrivingCars.

1. INTRODUCTION

One of the applications of Machine Learning ML and Deep Learning DL is in the fieldof autonomous driving or self-driving cars. It is a new high technology that might op- erate the self-driving of future cars. As a baseline algorithm, CNN (Convolutional Neu-ral Network) model is used to predict the control command from the video frames. One interesting task of this control system is to recognize different traffic signs present on the road to guaranty safe driving [1]. For this purpose, a CNN model is trained on mappixels from processed images taken from cameras and sensors placed on the car. This kind of model proved its performance in many other works such as medical imaging, pattern recognition (text, speech, etc), computer vision, and other interesting applica- tions [2]. Several benefits can be achieved using this high technology, notably: the reduction of deaths and injuries in road accidents, reduction of air pollution, increasingthe quality of car control, etc. in one word the main objective is to achieve the safety ofhumans. In this way, an automatic system of detection helps the driver to recognize thedifferent signs quickly and consequently some risks, especially when this driver is in abad mental state or drives his car in a crowded city or any other complex environment,which can cause the driver to overlook messages sent from the traffic signs put on the side of the road. Thus, the sign is to report correct messages as soon as possible to the driver and then reduce the burden of the driver and increase the safety of driving and decrease the risk of accidents. The present paper is organized as follows: section 1 is ashort introduction presenting the area of our work and its advantages and benefits. Sec-tion 2 is a detailed overview of the related works in the same area. In the third section, we described our proposed model. The fourth section presents the experimental part wehave done to validate our proposed model. In section 5, we illustrated the obtained results when applying our new model. In section 6, we discussed the resultsobtained in the previous section. In the last section, we summarized the realized work and sug- gested some perspectives for future researches.

2. RELATED WORK

The idea of autonomous driving started at the end of the 1920s, but the first autonomouscar appeared in the 1980s. some promising projects have been realized in this period such as the autonomous car called NAVLAB in 1988 and its control system ALVINN[3], [4]. Among the most important tasks in the field of self-driving cars or autonomousvehicles, we find the traffic signs recognition. For this purpose, several methods basedon feature extraction have been developed, including Scale-Invariant Feature Transfor-mation (SIFT) [5], Histogram of Oriented Gradient (HOG) [6], and Speed Up Robust Feature (SURF) [7]. The use of ANN in autonomous driving is not new, Pomerleau used in ALVINN system a fully connected neural network with a single hidden layer with 29 neurons to predict steering commands for the vehicle. The rise in machine learning and DL, especially, the famous CNN models helped to improve significantly the performance of the traffic signs detection and surprising results have been achieved[8]. In 2004, the company DARPA seeded a project named DAVE or DARPA autono-mous vehicle (Net-scale Technologies 2004) based on the use of the CNN model. Manyyears later, some new methods based on CNNs have been also developed, it is the caseof the method called semantic segmentation aware SSA [9] and the DP-KELM method[10], [11]. More recently, the Nvidia team trained

large CNN mapping images obtainedfrom driving a real car to steering commands [12]. Today, CNN models have been developed and applied on many other interesting applications, notably: AlexNet [13], VGG [14], GoogleNet [15], ResNet [16], R-CNN series [17], [18], Yolo [19], SDD [20], R-FCN [21]. These models are widely used by researchers in different areas of object recognition andgave an excellent performance on most of the available datasets. This encourages researchers in the field of traffic signs recognition and self-driving cars to develop new models more performant and accurate. For instance, GoogleNet struc- turewhich isa multilabel CNNneural network has been used for road scenerecognition [22]. Similarly, ResNet architecture based

3. THE PROPOSED MODEL

In the present work, we have developed an automatic system that allows us to detect and classify some given images representing traffic signs panels. For this purpose, we have proposed a CNN model composed of manyconvolutional layers, max-pooling layers, and a fully connected layer. As programming tools, we have used python, Ten-sorflow, and Keras which are the most used in the field. Fig 1 presents a detailed dia- gram of the proposed CNN model to improve the performance of the classification task.

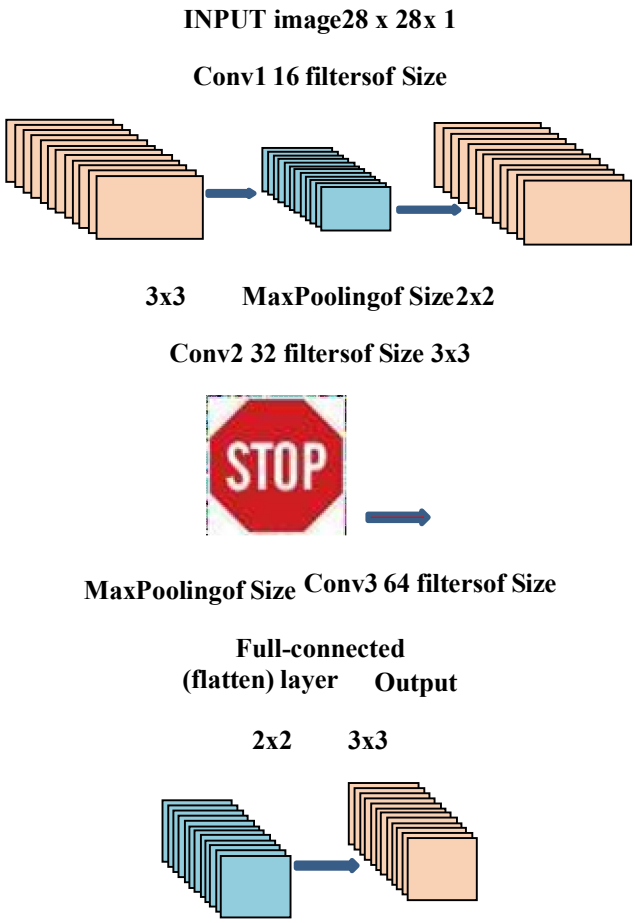


Fig .1: The architecture of the proposed CNN model

4. EXPERIMENTAL WORK

Used dataset

In our experiments, we have used the Belgium traffic signs dataset which is a collectionof images usually written in French or Dutch because these two languages are the offi- cial and the most spoken languages in Belgium. The collection canbe divided into six (06) categories of traffic signs: warning signs, priority signs, prohibitory signs,mandatory signs, parking and standing on the road signs, designatory signs. After downloading Belgium traffic signs files (training and testing files), we take a look at the folder structure of this data set, we can see that the training, as well as the testing data folders, contain 62 subfolders, which present 62 types of traffic signs used for classification. All images have the format (. ppm: Portable PixMap). Thus, the per- formed task is to classify a given image into one of 62 classes representing traffic signspanels. Fig 2 represents an illustration of some traffic sign panels issued of the Belgiumtraffic signs dataset, Fig 3 gives the distribution of these panels by class (type/group).

Fig .2. Examples of images in Belgium Traffic Signs Dataset

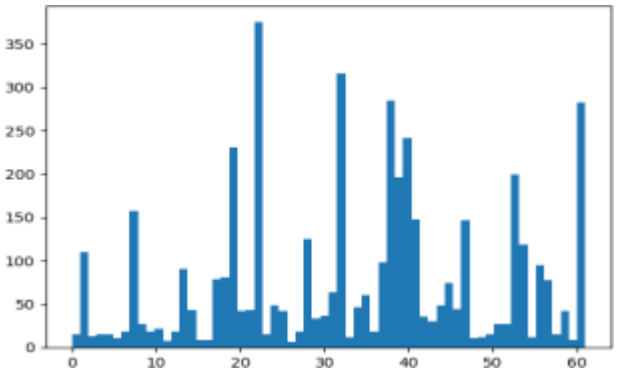


Fig .3. Distribution of panels by labels in Belgium Traffic Signs Dataset

Python: Python is currently one of the most popular languages for scientific applica- tions. It has a high-level interactive nature and a rich collection of scientific libraries which lets it a good choice for algorithmic development and exploratory data analysis.It is increasingly used in academic establishments and also in industry. It contains a famous module called the scikit-learn tool integrating a largenumber of ML algorithmsfor supervised and unsupervised problems such as decision trees, logistic regression, Naïve Bayes, KNN, ANN, etc. this package of algorithms allows to simplify ML to non-specialists working on a general-purpose.

Tensorflow: TensorFlow is a multipurpose open-source library for numerical compu-tation using data flow graphs. It offers APIs for beginners and experts to develop for desktop, mobile, web, and cloud. TensorFlow can be used from many programming languages such as Python, C++, Java, Scala, R, and Runs on a variety of platforms including Unix, Windows, iOS, Android. We note also that Tensorflow can be run on single machines (CPU, GPU, TPU) or distributed machines of many100s of GPU cards.

Keras: Keras is the official high-level API of TensorFlow which is characterized by many important characteristics: Minimalist, highly modular neural networks library written in Python, Capable of running on top of either TensorFlow or Theano, Large adoption in the industry and research community, Easy production of models, Supportsboth convolutional networks and recurrent networks and combinations of the two, Sup-ports arbitrary connectivity schemes (including multi-input and multi-output training),Runs seamlessly on CPU and GPU.

5. EVALUATION

To validate the different ML algorithms, and obtain the best model, we have used the cross-validation method consisting in splitting our dataset into 10 parts, train on 9 and test on 1, and repeat for all combinations of train/test splits. For the CNN model, we have used two parameters which are: loss value and accuracy metric.

- ***Accuracy metric:*** This is a ratio of the number of correctly predicted instances di- vided by the total number of instances in the dataset multiplied by 100 to give a per- centage (e.g., 90% accurate).
- ***Loss value:*** used to optimize an ML algorithm or DL model. It must be calculated on training and validation datasets. Its simple interpretation is based on how well the ML algorithm or the DL built model is doing in these two datasets. It gives the sum oferrors made for each example in the training or validation set.

6. ILLUSTRATION OF THE OBTAINED RESULTS

To build an efficient predictive model and achieve a higher accuracy rate, we have performed the following task:

Designing a CNN (Convolutional Neural Network) model composed of many layers asit was presented in section 3 and Fig.

We can also describe our proposed model as follows:

- The first convolutional layer Conv1 constituted of 16 filters of size (3x3).
- A Max-Pooling layer of size (2x2) allowing the reduction of dimensions (weigh, high) of images issued of the previous layer after applying the different filters of Conv1.
- A second convolutional layer Conv2 constituted of 32 filters of size (3x3).
- A Max-Pooling layer of size (2x2) allowing thereduction dimensions (weigh, high)of images issuedof the previous layer after applying the different filters of Conv2.
- A third convolutional layer Conv3 constituted of 64 filters of size (3x3).

A flatten Layer

- A full connected layer FC of size 100 allowing to transform the output of the previ-ous layer into a mono-dimensional vector.
- An output layer represented by a reduced mono-dimensional vector having as sizethe number oftraffic signs classes (62).
- For all the previous layers a "Relu" activation function and a "softmax" functionare used to normalize values obtained in each layer

Table 1. Description of the Proposed CNN Model

Layer Type	Output Shape	Nb. parameters
conv2d_1 (Conv2D)	(None, 26, 26, 16)	160
max_pooling2d_1 (MaxPooling2)	(None, 13, 13, 16)	0
conv2d_2 (Conv2D)	(None, 11, 11, 32)	4640
max_pooling2d_2 (MaxPooling2)	(None, 5, 5, 32)	0
conv2d_3 (Conv2D)	(None, 3, 3, 64)	18496
flatten_1 (Flatten)	(None,576)	0
dense_1 (Dense)	(None, 62)	35774
Total parameters		59,070
Trainable parameters		59,070
Non-trainable parameters		0

To validate our CNN model, we have used two parameters which are: loss value and accuracy metric. Below pseudocodewritten in Tensorflow and Keras which allowed usto build our model.

```
model = Sequential()
model.add(Convolution2D(16,(3,3),activation=
'relu',kernel_initializer='he_uniform', input_shape=(28,28,1)))
model.add(MaxPooling2D(2, 2))

model.add(Convolution2D(32, (3,3), activation='relu',
kernel_initializer='he_uniform'))
```

model.add(MaxPooling2D(2, 2))

model.add(Convolution2D(64, (3,3), activation='relu', kernel_initializer='he_uniform'))

model.add(Dropout(0.25)) model.add(Flatten()) model.add(Dense(100,activation='relu',

kernel_initializer='he_uniform')) model.add(Dense(62, activation='softmax'))

Tables 2 below summarize the obtained results after applying the proposed CNN model.

Table 2. Loss value and accuracy value obtained when applying the proposed model

	Loss value	Accuracy value
Training set	0.0921	97,31%
Test set	0.0169	99,80%

Fig .5: Training loss Vs Validation loss of the CNN model

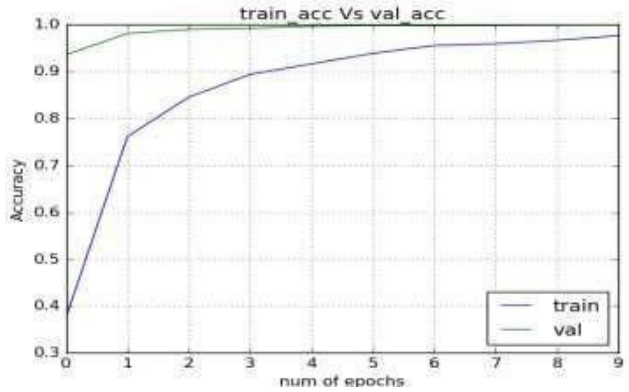


Fig .6. Training accuracy Vs Validation accuracy of the CNN model

7. DISCUSSION

Table 2 presents the obtained results when applying the proposed CNN model on thetraining set and the test set. Two performance measures are considered in this case, theloss value which calculates the sum of errors after training the model, and the accuracyvalue which gives the rate of correctness. It seems clear that the loss value is very lowagainst the accuracy which is very high and depends on the size of the used set. It is thereason for which the accuracy of the training set is higher than the accuracy of the testset (in our case they are very closest).

In the same way, Fig 5 shows the evaluation of training loss and validation loss over time in terms of the number of epochs. It begins very high for the training and the test sets and ends very low when increasing the number of epochs.

Similarly, Fig 6 plots the evolution of training accuracy and validation accuracy in terms of the number of epochs. Contraryto the loss value, the accuracy starts very lowand ends very high. This property is clearer with the training set because of its large size.

Finally, we can also underline that the performance of our proposed model (the classi-fication accuracy) is veryhigh (99.80%) compared to other realized models cited in theliterature section (section 2) which helps to increase the quality of car control by rec- ognizing all traffic signs placed on the road, and consequently, to guaranty more safetyfor humans and vehicles. We can extend this work to other object detection such as pedestrians, animals, and other complex obstacles.

8. CONCLUSION

In this paper, we propose an efficient traffic sign recognition method based on a CNN model for self-driving cars. The proposed method achieves high accuracy and fast processing time, making it suitable for real-time traffic sign recognition in self-driving cars. Our future work includes testing the proposed method on real-world data and integrating it into a self-driving car platform.

9. REFERENCE

[1] D.A. Pomerleau. Alvinn: an autonomous land vehicle in a neural network. Technical report, Carnegie Mellon University, ComputerScience Department. (1989).

[2] B.T. Nassu, M. Ukai : Automatic recognition of railway signs using sift features. In Intelli-gent Vehicles Symposium, pages 348–354. (2010)

[3] I.M. Creusen, R. G. J . Wijnhoven, E. Herbschleb, and P. H. N. D. With: Color exploitationin hogbased traffic sign detection. In IEEE International Conference on Image Processing, pages 2669–2672. (2010)

[4] J. Duan, and M. Viktor: Real time road edges detection and road signs recognition. In In- ternational Conference on Control, Automation and Information Sciences, pages 107–112. (2015)

[5] Intelligent Computing and Optimization. Proceedings of the 3rd International Conferenceon IntelligentComputing and Optimization 2020 (ICO 2020). https://link.springer.com/book/10.1007/978-3-030-68154-8

[6] S. Gidaris, and N. Komodakis: Object detection via a multi-region and semantic segmenta-tion-aware cnnmodel. In IEEE International Conference on Computer Vision, pages 1134–1142. (2015)

[7] X. Zeng, W. Ouyang, and X. Wang: Multi-stage contextual deep learning for pedestrian detection. In IEEEInternational Conference on Computer Vision, pages 121–128. (2013)

❑❑❑

A Case Study on Metar Data Forecasting Using Time Series

GOKULAPRIYA V[1], RABINTHA J[2]
[1]Assistant Professor, [2]Student
[1,2]Department of CSE, Mahendra College of Engineering, Salem, Tamil Nadu, India

ABSTRACT

Forecast is a common data science task that helps organisations with capacity planning, goal setting and anomaly detection. We using Time Series algorithm to forecast the metar data. Time series data, also referred toas time-stamped data. These data points typically consist of successive measurements made from the same source over a fixed time interval and are used to track change over time. In addition, traditional time series models like ARIMAX, SARIMAX have many stringent data requirements like stationarity and equally spaced values can be highly complex and difficult to work. So, one of the time series algorithm ***Facebook Prophet*** *which giving quick, powerful, and accessible time-series modelling to data analysts and data scientist. The model can useit in the python although it can also be implemented.*

Index Terms - Metar Data Forecast, Face Prophet, Python, R.

1. INTRODUCTION

METAR stands for Meteorological Aerodrome Report. Visibility is an important factor in all phases of flight, but especially when the aircraft is manoeuvring on or close to the ground. Poor visibility at a destination can reduce capacity of airports leading to ground delays, flight diversions, flight cancellations and extra operating costs. Conventional methods used to predict or forecast the visibility of the weather involves complex meteorological processes which takes much long time to analyse the data and produce a prediction. Therefore, it is necessary to develop a quick and easy system for forecasting the weather by quicklyassessing the given data.

Facebook prophet is one of the quick accessing libraries developed by Facebook's Core Data Science team and is designed for forecasting time series data and it is available as open source.

We can forecast the visibility in the air for the upcoming hours using independent attributes such as temperature, relative humidity, wind direction, wind intensity and so on.

2. LITERATURE REVIEW

Meteorologists predict weather and climate, study, analyse, interpret, andprovide weather reports and weatherpatterns on a day-to-day basis.

Atmospheric science deals with the Earth's atmospheric conditions and itsphenomena, such as the precipitationof typhoons, tornadoes, thunderstorms, rain, and snow. It also deals with variations of temperature, moisture, wind speed, wind direction, and similar patterns that produce distinct weather conditions.

Meteorologists predict weather forecasts using complicated mathematical equations. They use an anemometer to measure the speed and pressure of the wind. Instrumentsalso include barometers (one of the most critical instruments in weather forecasting), rain gauges to measure rainfall, wind vanes to measure wind speed, thermometers to measure temperature, weather balloons, and weather satellites to compile the data required for forecasting.

At present, meteorologists prefer to use other tools to forecast the weather which include:[6]

- Doppler radar
- Satellite Data
- Radiosondes
- Automated surface-observingSystems
- Supercomputer
- Supercomputers
- AWIPS

Here, the prediction can be made using the past data of weather that would be forecast using the time series algorithm that can be efficient and make the meteorologist know the earth's atmospheric phenomena.

There are several interesting facts are mentioned in this paper as follows, sales forecasting based on real-world data were analysed by Emir Zunic et al. [2] using Facebook prophet and 25 years of time series forecast using traditional algorithm called ARIMA (Autoregressive Integrated Moving Averages) by Jan G De Gooijer et al. [3], and forecasting international quarterly tourist flows using error- correction and time-series models by N. kulendran et al.[4] that include error-correction model and the autoregressive model which based on time series model that predicts the quarterly tourist flows into Australia from the major tourist markets of USA, Japan, UK and New Zealand.

Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India were analysed by Priya Narayanan et al. [5] they analysis a period of 60 years of monthly Rainfall data for March, April, May (MAM) for six stations (Abu (Ab), Ahmedabad (Ah), Ajmer (Aj),

Amritsar (Am), Bikaner (Bk), Jodhpur (Jd). The magnitude and practical significance of trend has been estimated using the Theil and Sen's median slope estimator, and assessing the percentage change over the mean for the period concerned (Yue and Hashino,2003; Yue et al.,2002.

3. ALGORITHM

Time series analysis comprises methods for analysing time-series datain order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

Facebook Prophet is an open-source algorithm for generating time-series models that uses a few old ideas with some new twists. It is particularly good at modelling time series that have multiple seasonality's and doesn't face some of the above drawbacks of other algorithms. At its core is the sum of three functions of time plus an error.

The growth function has three mainoptions:

Linear Growth: This is the default setting for Prophet. It uses a set of piecewise linear equations with differing slopes between change

$$g(t) = \frac{C(t)}{1 + x^{-k(t-m)}}$$

points. When linear growth is used, the growth term will look similar to theclassic $y = mx + b$ from middle school, except the slope(m) and offset(b) are

Flat: Lastly, you can choose a flat trend when there is no growth over time (but there still may be seasonality). If set to flat the growthfunction will be a constant value.[1]

For example:
Consider a Metar dataset that includes Station, Timestamp, Air temperature, Dew point temperature, Relative humidity, Wind, direction, Wind speed, One hour precipitation, Pressure altimeter, Sea level pressure, Visibility, Wind gust, and so on term: growth g(t), seasonality s(t),holidays h(t) , and error e_t :[1]

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

variable and will change value ateach changepoint.

Logistic Growth: This setting is useful when your time series has a cap or a floor in which the values you are modelling becomes saturated and can'tsurpass a maximum or minimum value (think carrying capacity). When logistic growth is used, the growth term will look similar to a typical equation for a logistic curve (see below), except it thecarrying capacity (C) will vary as a function of time and the growth rate (k)and the offset(m) are variable and will change value at each change point.

Installing library **fbprophet** and weextracting the required columns. Then constructing heat map that used to show relationship between two variables.

We had Univariate and Multivariate forecasting of Metar data which of timeseries forecasting involves in findingthe future values.

The univariate time series has only one variable, a multivariate has more than two variables. After that as usual training and testing process takes place. Then, calculating the accuracy of tested data which are in the continuous data format

The prophet procedure is an additive regression model with four main components:

A piecewise linear or logistic growth curve trend. Prophet automatically detects changes in trends by selecting changepoints from the data. It includesa yearly seasonal component modelled using Fourier series and a weekly seasonal component modelled using dummy variables.

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data.

Takes dependent and independent variables to plot and know the yearly, weekly and daily seasonality with the help of Facebook prophet. In the traditional algorithm ARIMA the same can be done but it requires separate library. So, we going with the fbprophet to predict the weather by the meteorologist.

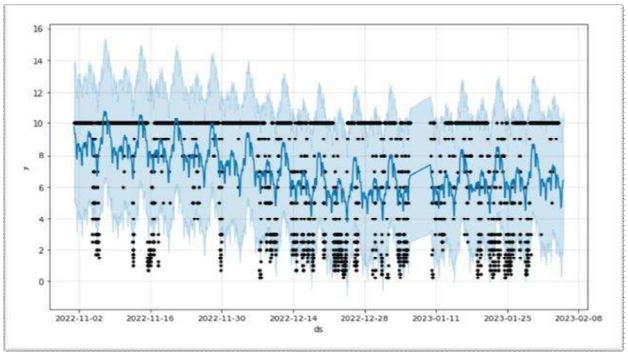


Fig.A.2. Visualizing the metar prediction for the next 24 hours

4. CONCLUSION

In the metar data forecasting we using Facebook prophet algorithm for easy and fast accessing although it doesn't face many of challenges that some other kinds of time-series modelling algorithm face.

The time series algorithms to analyse the pattern in METAR data and thereby forecasting the upcoming weather would significantly improve the efficiency of the

prediction.

5. FUTURE ENHANCEMENT

The traditional model like ARIMA and SARIMA both algorithms for forecasting.

ARIMA takes past data and predicts the future.in case of SARIMA similarly takes past data and also takes any seasonality patterns.

But compare with another model FACEBOOK PROPHET model is **interpretability.**

A person without prior experience or in-depth knowledge in time series modelling can work around. Facebook's prophet is accurate and fast.so, the fbprophet is used for the metar forecasting aviation weather is crucial for ensuring the safe and efficient operation on flights.

Accurate weather forecasting helps pilots make informed decision about whether to fly or not and helps them plan the most efficient and safe route,

It also helps in minimizing the impact of weather on passenger comfort and helps ensure that passengers arrive at their destination on time and is good condition.

6. REFERENCES

[1] Time series analysis with Facebook Prophet using the covid-19 data by Mitchell Krieger at Feb 20, 2021. https://towardsdatascience.com/time- series-analysis-with-facebook-prophet- how-it-works-and-how-to-use-it-f15ecf2c0e3a

[2] Emir Zunic, Kemal Korjenic, Kerim Hodzic and Dzenana Donko. Prediction on sales forecasting based on Real-World data. International Journal of Computer Science & Information Technology (IJCSIT) Vol 12, No 2, April 2020

[3] Jan G De Gooijer, Rob J Hyndman. 25 years of time series forecasting. Journal of forecasting 1982-1985; International Journal ofForecasting 1985-2005.

[4] Kulendran, N., & King, M.L. (1997). Forecasting international quarterly tourist flows using error- correction and time-series models. International Journal ofForecasting, 13, 319–327.

[5] Priya Narayanan, Ashoke Basistha, Sumana Sarkar, Kamna Sachdeva, Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India, C. R. Geoscience. 345 (2013) 22–27.

[6] meteorologist uses some tools to forecast the weather refer the link. https://www.noaa.gov/stories/6-tools- our-meteorologists-use-to-forecast-#:~:text=Supercomputers&text=Observational%20data%20collected%20by%

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Artificial Intelligence in Dairy Farming

V. MANIBABU[1], DR. M. GOMATHY[2], DR. V. JAYALALITHA[3]
[1,2]Research Scholar, [3]Assistant Professor
[1,2,3]Department of Computer Science and Engineering,
[1,2]Shrimati Indira Gandhi College, Trichy, Tamil Nadu, India
[3]Veterinary University Training and Research Centre, Trichy Tamil Nadu, India.

ABSTRACT

The word Artificial intelligence (AI) is nothing but the meaning of how the machine learns our thoughts and ideas converting into Digital computer science. The AI can also make decisions and carry out actions based on our ideas and suggestions etc., on behalf of human being wisdoms. AI has multiple technologies i.e., software, hardware component which helps machining learning. In our review report, we have to explain how AI can help in our dairy farming activities. For eg. AI can detect Cow's pain, Diseases affects Cow's limbs due to infection or any injury, wound etc. It can also help to improve the quality of milk yields, Improve Milk Protein, Milk Fat and concentrate feed intake. All these analyses can be possible by using statistical methods or predictive analytics in dairy farm, but AI methods do extraordinary performance beyond our expectations.

Index Terms - Algorithms, Machine Learning.

1. INTRODUCTION

The use of AI in dairy farming is becoming increasingly popular, with several applications being developed to modernize the industry. Technologies such as predictive software, robot farmhands, and data-capturing drones are being developed to help monitor cow health and welfare, detect early-stage lameness in cattle, and provide performance indicators for cows and farms. AI can also be used for breeding recommendations and to provide detailed information on dairy products for consumers. Researchers believe that AI will eventually provide broader services to help protect cow health, boost milk production, and improve farm productivity.

Artificial intelligence will play a major role in near future of dairy industry. If there is an open source data for dairy farming it will be much useful for the small scale dairy farmers. It will be helpful in policy makers. India stands first position in milk production among worldwide due to its 100 million number of dairy farmers having 1-2 dairy cows.

This review is dealt about applications of artificial intelligence in dairy farms and scope of AI in dairy business. Dairy farming is vast industry where there is lot of opportunities for multiple analyses, computation and decision making. Monitoring the dairy farming like watching herd at grazing, identifying the sick animals and in heat, feed formulation, effective managemental practices and treatment protocols and their responses, milk production records etc. Descriptive analysis based on the data of previous records and predictive analysis to understand the performance through statistical modeling, for example estimating which cows are at risk of disease given their historical data.

AI support predictive analytics in dairy farm. Conventionally, statistical methods may perform this prediction, but AI methods do extraordinary performance beyond our expectations. There are three trends of AI methods in dairy farms and the first is about the volume, frequency, and variety of data collected routinely in dairy farms, such as by precision dairy technologies and genomic testing. Precision dairy technology is the continuous monitoring of animals' behavior, milk constituents, milk yield, video analysis, record analysis, and physiological monitoring in real time (Eckelkamp, 2019). The detection of disease, estrus, or both is a common goal. According to Pugliese et al., 2021, decision trees and convoluted neural networks, may be applied to a wide range of problems, including visual recognition, speech recognition, and natural language processing and this may be the second trend. The development of online and edge computing power, which enables the timely execution of cutting-edge AI algorithms, is a third trend (Shalf, 2020).

2. STUDIES RELATED TO DAIRY FARMING AND ARTIFICIAL INTELLIGENCE

Most of the studies focused problems related to the physiology and health of dairy cows (32%), and feature data (explanatory or independent variables) were most often derived from sensors (48%). The largest number of studies employed tree-based algorithms (54%). Shine and Murphy (2021) also observed that from 2018 to 2021, there was more than a 7-fold increase in the number of studies that focused on the physiology and health of dairy cows. This compares to almost a 3-fold increase in the overall number of publications, suggesting an increased focus on this sub domain of health and physiology. In addition, a 5-fold increase in the number of publications that employed neural network algorithms (deep learning) was identified since 2018, in comparison with a 3-fold increase in the use of both tree-based algorithms and statistical regression algorithms, suggesting an increasing

use of neural network–based algorithms.

Researchers from Australia demonstrated for performing temperature regulator through AI in dairy farming. They proposed that the performace of individual cow can be assessed through a smartphone app. They reported that their model outputs.

Artificial intelligence (AI) application-based radio frequency identification system (RFID) for specific cow data input and machine learning (ML) processing (Source: Fuentes et al., 2020)

To estimate feed intake of individual cows, Martin et al. (2021) compared methods to predict individual feed intake and residual feed intake using data streams of behavioral, metabolite data and classical performance variables from cows with known individual feed intakes. Multiple linear regressions had analogous performance compared with Machine Learning models. Multiple data streams had more impact in model predictions compared with single data streams. Machine vision is another approach to calculate feed intake by comparing absence of feed in the feed trough of the cow. Machine vision is used to identify the cow as well as the feed that eaten. Saar et al. (2022), reported that cameras used to acquire images of feed composition and disappearance. They concluded that red-green-blue-depth cameras and the deep learning model have the potential to measure individual feed intake and could be tuned to different types of feed of dairy cows. Lassen et al. (2018) demostrated an implementation of a 3-dimentional camera for deep learning to measure individual feed intake in group-fed dairy cattle.

Robotics refers to robots built and programmed to do the very specific tasks. Laser guided detection of teat placement for attachment of the milking unit is accomplished by algorithms that appear intelligent. Other examples of robotics that appear so complex as to be considered intelligent are self-driving vehicles to deliver feed or work in crops.

3. CONCLUSION

Machine learning models may be used to assess welfare of animal and good quality production in terms of milk quality and quantity. Based on the data of the models may be developed for machine learning and can be applied to any dairy farm. Artificial Intelligence in dairy farms and the ML models developed may be adopted with limited technological facilities viz., automated gate, temperature regulator, automated feed provider etc. This study reviewed about the artificial intelligence and machine learning in various applications of dairy farm.

Conventional methods like regression models may perform better than AI and still there are some advantages. many studies describe AI applications is less practical by the user and the user have to make their software with the available data.

4. REFERENCES

[1] Eckelkamp, E. A. 2019. Invited Review: Current state of wearable precision dairy technologies in disease detection. Appl. Anim. Sci. 35:209–220. https://doi.org/10.15232/aas.2018-01801.

[2] Fuentes S, Gonzalez Viejo C, Cullen B, Tongson E, Chauhan SS, Dunshea FR. Artificial Intelligence Applied to a Robotic Dairy Farm to Model Milk Productivity and Quality based on Cow Data and Daily Environmental Parameters. Sensors (Basel). 2020 May 24;20(10):2975. doi: 10.3390/s20102975. PMID: 32456339; PMCID: PMC7285505.

[3] Lassen, J., J. R. Thomasen, R. H. Hansen, G. G. B. Nielsen, E. Olsen, P. R. B. Stentebjerg, N. W. Hansen, and S. Borchersen. 2018. Individual measure of feed intake on in-house commercial dairy cattle using 3D camera system. Pages 635–640 in Proc. 11th World Congr. Genetics Appl. Livest. Prod. Al Rae Center for Genetics and Breeding, Massey University. http://www.wcgalp.org/proceedings/2018.

[4] Martin, M. J., J. R. R. Dórea, M. R. Borchers, R. L. Wallace, S. J. Bertics, S. K. DeNise, K. A. Weigel, and H. M. White. 2021. Comparison of methods to predict feed intake and residual feed intake using behavioral and metabolite data in addition to classical performance variables. J. Dairy Sci. 104:8765–8782. https://doi.org/10.3168/jds .2020-20051.

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Secure Data Group Sharing and Dissemination on the Public Cloud with Attribute and Time Conditions

MAILSAMY M[1], MANJU B[2], HARISH T[3], PANDIYARAJA[4]
[1]Associate Professor, [2,3,4]Student
[1,2,3,4]Department of Computer Science and Engineering,
[1,2,3,4]Annapoorna Engineering College, Salem, Tamil Nadu, India

ABSTRACT

Users and companies all around the world are becoming more and more interested in cloud computing. Although users of public clouds can have their data protected using cryptographic approaches, a number of difficulties still need to be resolved, including the secure distribution of data groups and the fine- grained access control of time-sensitive data. In this research, we present an identity-based data group sharing and dissemination scheme in public cloud, where the owner could broadcast encrypted data to a group of receivers at once by identifying these receivers in a practical and secure manner. We use attribute- based and timed- release conditional proxy re-encryption to ensure that only data disseminators whose attributes satisfy the access policy of encrypted data can disseminate it to other groups after the releasing time by delegating a re- encryption key to cloud server. This approach enables secure and flexible data group dissemination. The accompanying attributes and releasing time of the re-encryption conditions enable the data owner to impose timed-release access control with fine-grained control over distributed ciphertexts. Theoretical analysis and experimental findings demonstrate that the computational overhead and expressive dissemination conditions are traded off in the suggested strategy.

Index Terms - ***.

1. INTRODUCTION

The popularity of cloud computing is obtained from the benefits of rich storage resources and instant access [1]. It aggregates the resources of computing infrastructure, and then provides on-demandservices over the Internet. Many famous companies are now providing public cloud services, such as Amazon, Google, Alibaba.

These services allow individual users and enterprise users to upload data (e.g., photos, videos and documents) to cloud service provider (CSP), for the purpose of accessing the data at any time anywhere and sharing the data with others. In order to protect theprivacy of users, most cloud services achieve access control by maintaining access control list (ACL). Inthisway, users can choose to either publish their data to anyone or grant access rights merely to their approvedpeople. However, the security risks have raised concerns in people, due to the datais stored in plaintext form by the CSP. Once the data is posted to the CSP, it is out of the data owner's control [2]. Unfortunately, the CSP is usually a semi- trusted server which honestly follows thedesignated protocol, but might collect the users' data and even use them for benefits without users' consents. On the other hand, the data hastremendous usages by various data consumers to learn the behavior of users.

In this way, data owner can customize fine- grained dissemination condition for the shared data.

2. OVERVIEW OF THE PROPOSED SYSTEM

INTRODUCTION:

The wellbeing of ladies as well as product conditions will be only a tick away atless expensive rate by machine and utilizing our normal framework The gadget will be setoff over the tapping button during crisis circumstance. A sectionphysically getting to the application this frenzy switch can likewise be utilized. During the frenzy circumstance the current area will be shipped off companions, family and furthermore to cops.

ARCHITECTURE OF THE PROPOSED SYSTEM:

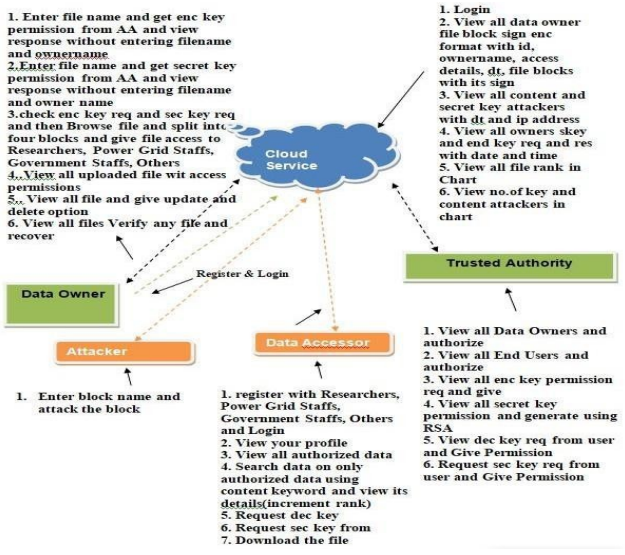


Figure: Architecture Diagram

MODULES:

Data Owners (DO) DO decide the access policy and encrypt thedata with CP-ABE. The encrypted data will be uploaded to the Cloud Servers. DO are assumed to be honest in the system.

Data Requester/Receivers (DR)

DR sends the decryption request to Cloud and obtain the ciphertexts over the internet. Only when their attributes satisfy the access policies of the ciphertext, can they get access to the plaintexts. Data requester/receivers may collude to access the data that is otherwise not accessible individually.

Cloud Servers (CS):CS are responsible for storing a massive volume of data. They cannot be trusted by DO. Hence, it is necessary for DO to define theaccess policy to ensure the data confidentiality. CS are assumed not tocollude with DR.

Trusted Authority (TA)

AA is responsible for registering users, evaluating their attributes and generating their secret key SK accordingly. It runs the Setup algorithm, and issues public key PK and master key MK to each DO. It is considered as fully trusted.

3. PROPOSED SYSTEM

The proposed system introduces asolutionto achieve cipher text group sharing among multiple users, and capture the core feature ofmultiparty authorization requirements. The contributions of our scheme are as follows:

- The system achieves fine-grained conditional dissemination over the cipher textin cloud computing with attribute based CPRE. The cipher text is firstly deployed with an initial access policy customized by data owner. Our proposed multiparty access control mechanismallows the data co-owners to append new access policies to the cipher text due to their privacy preferences. Hence, the cipher text can be re-encrypted by the data disseminator only if the attributes satisfy enough access policies.
- The system provides three strategies including full permit, owner priority and majority permit to solve the privacy conflicts problem. Specially, in full permit strategy, data disseminator must satisfy all the access policies defined by data owner and co- owners. With the majority permit strategy, data owner can firstly choose a threshold value for data co-owners, and the cipher text can be disseminated if and only if the sum of the access policies satisfied by data disseminator's attributes is greater than or equal to this fixed threshold.
- The system proves the correctness of our scheme, and conduct experiments toevaluate the performance at each phase to indicate theeffectiveness of our scheme.

Advantages

- The Data security is more since data co- owners can renew the cipher texts by appending their access policies as the dissemination conditions.
- The system is more secured due toContinuous policy enforcement in whichthedata owner's access policy is enforced in the initial cipher text as well as the renewedcipher text.

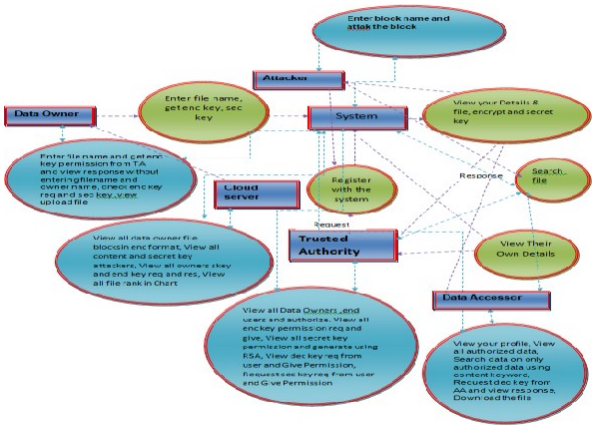


Figure: Data Flow Block Diagram

4. SYSTEM TESTING

TESTING METHODOLOGIES:

The following are the Testing Methodologies:

- Unit Testing.
- Integration Testing.
- User Acceptance Testing.
- Output Testing.
- Validation Testing.

The following are the types of Integration Testing:

Top-down Integration: This method is an incremental approach to theconstruction of program structure. Modules are integrated by movingdownward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

Bottom-up Integration

This method begins the construction and testing with the modules at the lowest level in the programstructure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom-up integration strategy may be implemented with the following steps:

- The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
- A driver (i.e.) the control program for testing is written to coordinate test case input and output.
- The cluster is tested.

- Drivers are removed and clusters are combined moving upward in the programstructure
- The bottom-up approaches test each module individually and then each module is integrated with a main moduleand tested for functionality.

User Acceptance Testing
User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

Output Testing
After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

Validation Checking:
Validation checks are performed on thefollowing fields.

Text Field:
The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in sometables and alphabetic in other tables. Incorrect entry always flashesand error message.

5. CONCLUSION

In this paper, we proposed a secure data group sharing and dissemination system on the public cloud with attribute and time conditions. The proposed system ensures that only authorized users can access the data and only during specified time periods. The system uses attribute-based encryption and time-based access control to provide a secure and efficient solution for data sharing and dissemination on the public cloud. The security and efficiency analysis showed that the proposed system is secure and scalable, making it suitable for various applications, including healthcare, finance, and government.

6. REFERENCES

[1] +Z. Yan, X. Li, M. Wang, and A. V.Vasilios, "Flexible data access controlbased on trust and reputation in cloud computing," *IEEE Transactions on CloudComputing*, vol. 5, no. 3, pp. 485-498, 2017.
[2] B. Lang, J. Wang, and Y. Liu, "Achieving flexible and self-contained data protection in cloud computing," *IEEE Access*, vol. 5, pp. 1510- 1523, 2017.
[3] Q. Zhang, L. T. Yang, and Z. Chen, "Privacy preserving deep computation model on cloud for big data feature learning," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1351-1362, 2016.
[4] H. Cui, X. Yi, and S. Nepal, "Achieving scalable access control over encrypted data for edge computing networks," *IEEE Access*, vol. 6, pp. 30049–30059, 2018.
[5] K. Xue, W. Chen, W. Li, J. Hong, and P.Hong, "Combining data owner-side and cloud-side access control for encrypted cloud

❑❑❑

Systematic Mapping: Artificial Intelligence Techniques in Software Engineering

DR. K. VELUSAMY[1], KOUSALYA[2]
[1]Professor and Head, [2]Student
[1,2]Department of Mechanical Engineering,
Annai Mathammal Sheela Engineering College, Erumaipatti, Tamil Nadu, India

ABSTRACT

Artificial Intelligence (AI) has become a core feature of today's real-world applications, making it a trending topic within the software engineering (SE) community. The rise in the availability of AI techniques encompasses the capability to make rapid, automated, impactful decisions and predictions, leading to the adoption of AI techniques in SE. With industry revolution 4.0, the role of software engineering has become critical for developing productive, efficient, and quality software. Thus, there is a major need for AI techniques to be applied to enhance and improve the critical activities within the software engineering phases. Software is developed through intelligent software engineering phases. This paper concerns a systematic mapping study that aimed to characterize the publication landscape of AI techniques in software engineering. Gaps are identified and discussed by mapping these AI techniques against the SE phases to which they contributed. Many systematic mapping review papers have been produced only for a specific AI technique or a specific SE phase or activity. Hence, to our best of knowledge within the last decade, there is no systematic mapping review that has fully explored the overall trends in AI techniques and their application to all SE phases.

Index Terms - ***.

1. INTRODUCTION

Software development process is a very complex process that, at present, is primarily a human activity. Programming in software development, requires the use of different types of knowledge: about the problem domain and the programming domain. It also requires many different steps in combining these types of knowledge into one final solution. There are various techniques in artificial intelligence (AI) from the standpoint of their application in software engineering that can be deployed in solving problems associated with software development processes. Artificial Intelligence is concerned with the study and creation of computer systems that exhibit some form of intelligence and attempts to apply such knowledge to the design of computer-based systems that can understand a natural language or understanding of natural intelligence. Many Software products costs can be attributed to the ineffectiveness of current techniques for managing this knowledge, and Artificial Intelligence techniques can help alleviate this situation.

Software engineering and artificial intelligence are the two important fields of the computer science. Artificial Intelligence is about making machines intelligent, while Software engineering is knowledge –intensive activity, requiring extensive knowledge of the application domain and of the target software itself. This study intends to review the techniques developed in artificial intelligence from the standpoint of their application in software engineering. The goal of this research paper is to give some guidelines to use the artificial intelligence techniques that can be applied in solving problems associated with software engineering processes. The aim of this paper is to find out the exact AI technique is likely to be fruitful for particular software development process.

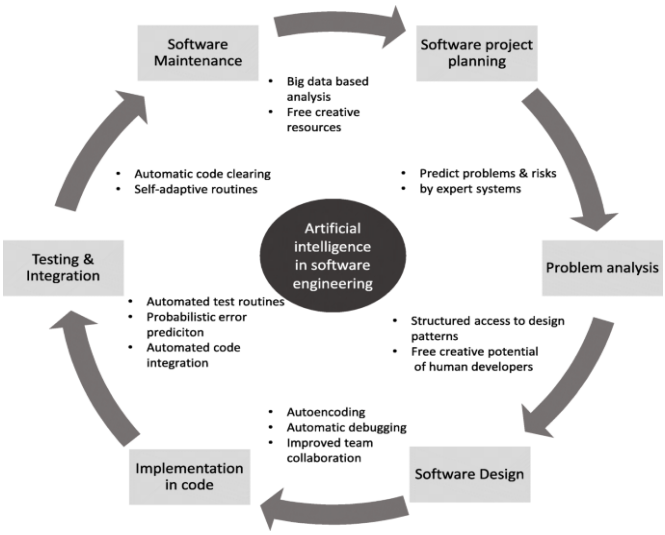


Fig 1: AI With Software Engineering

Because of the evolutionary nature of software products, by the time coding is completed, requirements would have changed (because of the long processes and stages of development required in software engineering): a situation that results in delay between requirement specification and product delivery. There is therefore a need for design by experimentation, the feasibility of which lies in automated programming. Some of the techniques and tools that have been successfully demonstrated in automated programming environments include: • Language Feature: this technique adopts the concept of late binding (i.e., making data structures very flexible). In late binding, data

structures are not finalized into particular implementation structures. Thus, quick prototypes are created which result in efficient codes that can be easily changed. Another important language feature is the packaging of data and procedures together in an object, thus giving rise to object-oriented programming: a notion that has been found useful in environments where codes, data structures and concepts are constantly changing. Lisp provides these facilities.

Artificial Intelligence (AI) techniques have been successfully applied in many areas of software engineering. The complexity of software systems has limited the application of AI techniques in many real world applications. This talk provides an insight into applications of AI techniques in software engineering and how innovative application of AI can assist in achieving ever competitive and firm schedules for software development projects as well as Information Technology (IT) management. The pros and cons of using AI techniques are investigated and specifically the application of AI in IT management, software application development and software security is considered.

Organisations that build software applications do so in an environment characterised by limited resources, increased pressure to reduce cost and development schedules. Organisations demand to build software applications adequately and quickly. One approach to achieve this is to use automated software development tools from the very initial stage of software design up to the software testing and installation. Considering software testing as an example, automated software systems can assist in most software testing phases.

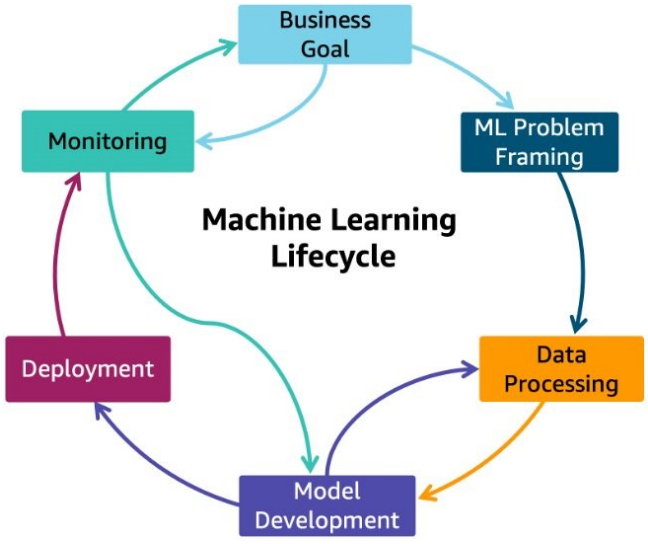


Fig 2: Machine Learning Lifecycle

On the hand data security, availability, privacy and integrity are very important issues in the success of a business operation. Data security and privacy policies in business are governed by business requirements and government regulations. AI can also assist in software security, privacy and reliability.

2. SOFTWARE REQUIREMENTS ANALYSIS

Requirement Engineering (RE):

Requirements are first expressed in natural language within a set of documents. These documents usually represent "the unresolved views of a group of individuals and will, in most cases be fragmentary, inconsistent, contradictory, not prioritized and often be overstated, beyond actual needs". The main activities of this phase are requirements elicitation, gathering and analysis and their transformation into a less ambiguous representation.

Knowledge Based Systems (KBS):

"The reuse of experts design knowledge can play a significant role in improving the quality and efficiency of the software development process". KBS were used to store design families, upon the development of the requirements, input and outputs of the system's functionality. The system searches the KB and proposes a design schema which is refined by the user to fully satisfy the requirements.

3. SOFTWARE ARCHITECTURE DESIGN

One of the most important problems facing the software engineer is to develop quality architecture from the requirements model. In this section we describe recent work on software architecture design using AI techniques. Developing the software architecture starts by defining a hierarchy of subsystems and components with allocated responsibilities from the information provided by the requirements and analysis models.

Software Coding and Testing:

Techniques learned from AI research make advanced programming much simpler, especially with regard to information flow and control as a result of advances in knowledge representation. In the following we focus on the AI techniques used in supporting the tasks of coding and testing.

4. SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS

System	:	Pentium IV 2.4 GHz.
Hard Disk	:	500 GB.
Monitor	:	15 VGA Colour.
Mouse	:	Logitech.
RAM	:	4 GB.

SOFTWARE REQUIREMENTS

Operating System:		Windows-7/10(64-bit).
Language	:	Python 3.7
IDE Tools	:	Anaconda 3.0

5. TENSORFLOW

Introduction to TensorFlow

Introduction to TensorFlowTensorFlow is a multipurpose open source so2ware library for numerical computation using data flow graphs. It has been designed with deep learning in mind but it is applicable to a much wider range of problems. In this tutorial I will cover the very basics of

TensorFlow not going much into deep learning

Source Code

```
#!/usr/bin/env python
# coding: utf-8
# In[42]:
import math
import pandas as pd
importnumpy as np
importmatplotlib.pyplot as plt
importseaborn as sns
importjoblib
importsklearn
importyaml
import sys
import glob
importos
get_ipython().run_line_magic('matplotli
b', 'inline')
fromsklearn.model_selection      import
KFold
fromsklearn.model_selection      import
train_test_split
fromsklearn.linear_model         import
LinearRegression
fromsklearn.ensemble             import
GradientBoostingRegressor
fromsklearn.preprocessing        import
MinMaxScaler
fromsklearn.metrics              import
mean_squared_error
get_ipython().run_line_magic('matplotli
b', 'inline')
sns.set(font_scale=1.2)
sns.set_style("white")
# In[43]:
data = pd.read_csv('fpgadata.csv')
#print(data.head(10))
print (data)
# In[44]:
data.columns
# In[45]:
print("\n\nData Information\n")
print(data.info())
# In[46]:
data.isnull().sum()
# In[47]:
print("\n\nMissing Data\n")
sns.heatmap(data.isnull(),yticklabels=F
alse,cbar=False,
cmap='rainbow')

# In[48]:
data.shape
# In[49]:
data.isnull().sum()
# In[50]:
data.describe()
# In[51]:
print("\n\nDistribution     Plot     of
Dataset\n")
data['APP
ID'].hist(bins=30,color='green',alpha=0
.7)
# In[52]:
print("\n\nCountplot of Category.\n")
sns.countplot(x='CAT
ID',hue='CATEGORY',data=data,palette='r
ainbow')
# In[53]:
data['CATEGORY'].value_counts()
# In[54]:
data['CATEGORY'].value_counts().plot(ki
nd='pie')
# In[55]:


defsum_cpu(node_res):
 """Return   sum   of   CPU   resources
allocated to all nodes"""
cpu = sum([v['cpu'] for v in node_res])
returncpu
defread_placement(placement,   df_data,
flow_dr=250):
"""Read placement dict and write it to
df_data. Then return."""
df_data['num_flows'].append(placement['
input']['num_flows'])
df_data['num_sources'].append(placement
['input']['num_sources'])
df_data['source_dr'].append(placement['
input']['num_flows'] * flow_dr)

df_data['num_instances'].append(placeme
nt['metrics']['num_instances'])

df_data['max_e2e_delay'].append(placeme
nt['metrics']['max_endToEnd_delay'])

df_data['total_delay'].append(placement
['metrics']['total_delay'])
df_data['runtime'].append(placement['me
trics']['runtime'])

df_data['total_cpu'].append(sum_cpu(pla
cement['placement']['alloc_node_res']))
returndf_data

defread_results(results):
 """Read   result   files   matching   the
pattern and return df containing their
metrics"""
data    =    {'num_sources'    :    [],
'num_flows':   [],   'source_dr':   [],
'num_instances': [],
 'max_e2e_delay':   [],   'total_delay':
[], 'runtime': [], 'total_cpu': []}

 # iterate through result files
for res in glob.glob(results):
 # open and save metrics of interest
with open(res, 'r') as f:
```

```
placement = yaml.load(f, Loader=yaml.SafeLoader)
data = read_placement(placement, data)

returnpd.DataFrame(data).sort_values(by=['num_flows'])


# In[56]:


# read results
dataset = 'web_data'
sources = 'three_source_dr250'
results = f'placement_data/{dataset}/{sources}/'

df_true = read_results(results + 'true/*.yaml')
df_fixed = read_results(results + 'fixed/*.yaml')
df_linear = read_results(results + 'linear/*.yaml')
df_boost = read_results(results + 'boosting/*.yaml')
df_svr = read_results(results + 'svr/*.yaml')
df_ml = read_results(results + 'ml/*.yaml')


# In[57]:


#df_linear.head(10)


# In[58]:


def plot(x_col, x_label, y_col, y_label, save_plot=True, plot_fixed=True):
sns.set(font_scale=1.3, style='white')
fig, ax = plt.subplots()

# plt.plot(df_true[x_col], df_true[y_col], label='True', color='black', marker='o')
ifplot_fixed:
plt.plot(df_fixed[x_col], df_fixed[y_col], label='Fixed', color='green', marker='+')
plt.plot(df_linear[x_col], df_linear[y_col], label='Linear', color='blue', marker='x')
# plt.plot(df_boost[x_col], df_boost[y_col], label='Boosting', color='red', marker='^')
# plt.plot(df_svr[x_col], df_svr[y_col], label='SVR', color='orange', marker='v')
plt.plot(df_ml[x_col], df_ml[y_col], label='SVR+Boosting', color='red', marker='s')

plt.xlabel(x_label)
plt.ylabel(y_label)
plt.legend()
```

6. REFERENCES

[1] H. K. Dam, ``Articial intelligence for software engineering," XRDS, Crossroads, ACM Mag. Students, vol. 25, no. 3, pp. 3437, Apr. 2019, doi:10.1145/3313117.

[2] D. P. Wangoo, ``Articial intelligence techniques in software engineering for automated software reuse and design," in Proc. 4th Int. Conf. Comput. Commun. Autom. (ICCCA), Dec. 2018, pp. 14, doi: 10.1109/CCAA.2018.8777584.

[3] A. Ahmad, C. Feng, M. Khan, A. Khan, A. Ullah, S. Nazir, and A. Tahir, ``A systematic literature review on using machine learning algorithms for software requirements identication on stack overow," Secur. Commun. Netw., vol. 2020, pp. 119, Jul. 2020, doi: 10.1155/2020/8830683.

[4] H. Alsolai and M. Roper, ``A systematic literature review of machine learning techniques for software maintainability prediction," Inf. Softw. Technol., vol. 119, Mar. 2020, Art. no. 106214, doi: 10.1016/j.infsof.2019.106214.

[5] N. Li, M. Shepperd, and Y. Guo, ``A systematic review of unsupervised learning techniques for software defect prediction," Inf. Softw. Technol., vol. 122, Jun. 2020, Art. no. 106287, doi: 10.1016/j.infsof.2020.106287.

[6] R. Malhotra, ``A systematic review of machine learning techniques for software fault prediction," Appl. Soft Comput., vol. 27, pp. 504518, Feb. 2015, doi: 10.1016/j.asoc.2014.11.023.

[7] S. K. Pandey, R. B. Mishra, and A. K. Tripathi, ``Machine learning based methods for

❑❑❑

A Survey on Textblob and Vader: Rule Based Model for Sentimental Analysis of IMDB Data

SENTHURAN S[1], METTUPATTI[2]
[1]Assistant Professor, [2]Student
[1,2]Department of Computer Science and Engineering,
Mahendra College Engineering, Salem, Tamil Nadu, India

ABSTRACT

The wide scope of the sentimental analysis is not the step to take in the NLP world, however, the good thing rule-based sentiment analysis requires minimal pre-works. Now we are using the TextBlob and VADER technologies to predict the accuracy of the data. Rule based is the approach of analyzing without text and training or using machine learning algorithm models. This commonly known as Lexicon based approach. Here we are implementing the natural language processing using logistic regression to predict the categorial data. Finally getting the high accuracy on text blob and Vader python libraries.

Index Terms - Rule Based Approach, Lexicon Based Model, TextBlob, Vader, NLP, NLTK and Data Pre-processing.

1. INTRODUCTION

Sentiment Analysis is a process that has analyse the emotion of a person, whether he happy or sad and analyse the positive negative of the command and reviews then, lot of scope and application into recommendation systems. Be a movie reviews, stock market, product, sentiments play a big role in analysing the trend and future product or service. Sentiment analysis can be turned into text is expressing positive, negative, or neutral sentiment. The paper mainly focuses on two approaches that are AI and Rule based reasoning. In python lot of packages are there for analyse the algorithm, now in this paper we are using two methods:

- Text Blob.
- VADER.

TEXTBLOB:

Textblob sentiment analyser is a python library used for Natural language processing that works on classification, translation and API keys.

For lexicon-based approaches a sentiment is a semantic orientation and intensity of the sentences. Textblob returns polarity and subjectivity of a sentence.

subjectivity by looking at the 'intensity'.itis a float that lies between [-1,1], -1 indicates negative and +1 indicates positive

$$R_{emp}(g) = \frac{1}{N} \sum_{i} L(y_i, g(x_i)).$$

text1 = TextBlob ("Today is a great day, but it is very boring"):

VADER:

VADER (for Valence Aware Dictionary for Sentiment Reasoning). We use combination of qualitative and quantitative methods to produce, and then empirically validate, a gold- standard sentiment lexicon that is microblog like contexts [1]. VADER not only tells the lexicon is positive, negative, or neutral, **it also tells how positive, negative, or neutral a sentence is.** The output from VADER comes in a Python **dictionary** in which we have **four keys** and their corresponding values. **'neg', 'neu', 'pos',** and **'compound'** which stands for Negative, Neutral, and Positive respectively [1]. The sum of pos, neg, neu intensities give 1. Compound ranges from -1 to 1 and is the metric used to draw the overall sentiment [2].

- positive if compound >=0.5
- neutral if -0.5 < compound
- < 0.5
- negative if -0.5 >=compound [2].

2. LITERATURE REVIEW

Sentimental analysis (also referred to as subjectivity analysis, opinion mining or emotional artificial intelligence) is natural language processing (NLP) technique that identifies important patterns of information and features of a large text corpus [4]. **Supervised learning** is the process of machine learning, trained in a labelled data which means predefined values. Input and outputs are well known to the user.It predicts the correctly with more accuracy because the machine already knows the result. It aims to predict the future data using the current input and output to find the exact correct answer.

Example: Fraud Detection, Spam Filtering.

3. ALGORITHM:

Logistic Regression in MachineLearning

- Logistic regression is a supervised learning algorithm used to predict a categorical data variable. If we have a largedata set then logistic regression may be used to predict the result.
- Logistic regression model is based on the logistic functions, which is a type of S-shaped curves for a

continuous input to a value of 0 and 1.
- Equation for logistic regression

→ p = 1/(1+e^(-q))
Where,
- P is the probability ofdependent variable.
- Q is the linear combination of independent coefficients.

The word logistic regressionrefers to the use of the logistic function to the relationship between the independent variables and the probability of event occurring.

4. PYTHON IMPLEMENTATION OF LOGISTIC REGRESSION (BINOMIAL):

Data set: IMDB MOVIE REVIEWDATASET
- The dataset comprises 50000 movie reviews from IMDB, of which 25000 are positive and 25000 negatives.

STEP 1: TOKENISATION
In this process, the given statement is broken up into separatewords.

E.g.: The movie was great!
- The
- Movie
- Was
- Great

STEP 2: DATA CLEANING
In this process, all the specialcharacters are removed.

E.g.:

STEP 3: STOP WORDS REMOVAL
In this process, all the words that add no value to the analysis part are removed.

E.g.: The, was removed.

STEP 4: CLASSIFICATION
In this process, all the remaining words are classified as positive, negative or neutral based on scores for each of the words in the dictionary, and a joint score is formed.

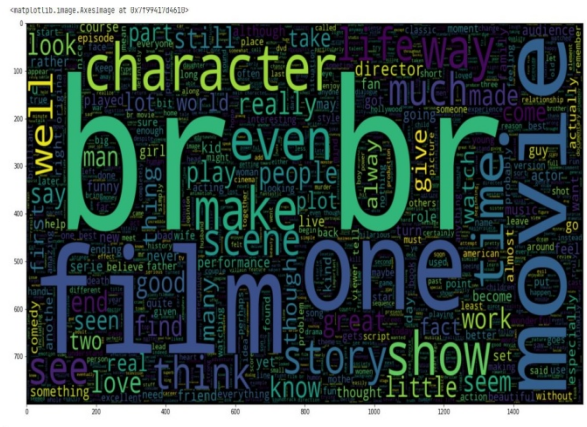
Eg: Classify the sentence:

Positive → +1
Negative → -1

The movie was great → positive - +1

Steps in Logistic Regression:
- Data Pre-processing step
- Fitting Logistic Regression tothe Training set
- Predicting the test result
- Test accuracy of the result

Train set: The training dataset is a set of data that was utilized to fit themodel. The dataset on which the model is trained. This data is seenand learned by the model.
- Test set: The test dataset is a subset of the training dataset that is utilized to give an accurate evaluation of afinal model Fit.[6]



Result for TextBlob:
- Out of 25000 positive reviews, TextBlob could detect 23681 as **positive with an accuracy of 94.724 %.**
- Out of 25000 negative reviews,TextBlob could detect 10749 as **negative with an accuracy of 42.996 %.**

Result for VADER:
- Out of 25000 positive reviews, Vader could detect 21523 **as positive with an accuracy of 86.092 %.**
- Out of 25000 negative reviews,TextBlob could detect 13274 as **negative with an accuracy of 53.096 %.**

5. CONCLUSION

The process of rule-based reasoning in TextBlob and VADER technologies, both the technologies predict more accuracy of correct results. Textblob predict the accuracy of positive reviews of 94%, whereas VADER predict the accuracy of negative reviews of 86%, as compared to both TextBlob predict the most heigh accuracy. The sentimental analysis using NLP is helpful to detect the dataset and produce the steps in data pre- processing, applying logistic regression easily predict the values of VADER data. So sentimental analysis plays a big role on the process of rule-based approach. The sentimental analysis helps to predict the data, it can be usedin different ways in future.

6. FUTURE ENHANCEMENT

Future work should implement Textblob and VADER for high accuracy and goodstructured models. For instance, the researchers can try to **remove the subjects in the sentences.** For example, recurrent neural network may be able to provide better performance. In this paper, compared both the TextBlob and Vader for sentimental analysis.
- **Methodology:** VADER andTextBlob are lexicon and rule-based technology.
- **Performance:** VADER is the fastest and best to

predict the result as compared to TextBlob is slower.

7. REFERENCES

[1] "VADER: A Parsimonious Rule- based Model for Sentiment Analysis of Social Media Text", Published on January 2015, At: Ann Arbor, MI **C.J. Hutto,** Georgia Institute ofTechnology

[2] "Rule-Based Sentiment Analysis in Python" Harika Bonthu — Published on June 18, 2021

[3] "Sentiment Analysis: VADER or TextBlob?" Afaf Athar — Published on January 6, 2021

[4] "Sentiment Analysis" Manika Lamba, Madhusudhan Margam, University of Delhi, Apíil 2022

[5] "Sentiment Analysis Without Modeling: TextBlob vs. VADER vs. Flair" Amy@GrabNGoInfo

[6] C. Kaur, A.Sharma, "Twitter Sentiment Analysis on Coronavirus using Textblob‖ in Proceedings of research gate", pp 1-12, 2020.

[7] S. Kiritchenko, X. Zhu, S. Mohammad, "Sentiment analysis of short informal Texts in Journal of Artificial Intelligence Research", Vol 50 No 11, pp 723-762, 2014.

[8] S. Tiwar and A. Sinha, "Sentiment Analysis of Facebook Data using Machine Learning in International Journal of Innovative Research in Applied Sciences and Engineering", Vol. 4, no 4, pp 735-742, 2020.

[9] "Sentiment Analysis of COVID- 19 Vaccine Rollout in India" Sushila Paliwal, Suraiya Parveen, M. Afshar Alam & Jawed Ahmed ,Conference paper First Online 04 January 2022

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Secure Data Sharing Using Cloud Through Blockchain for IOTEnvironment

S.JAYA PRAKASH[1], S.SARANYA DEVI[2], T.THENMOZHI[3], R.VENKATESH[4]
[1]Assistant Professor, [2,3,4]Student
[1,2,3,4]Department of Computer Science and Engineering,
Annapoorna College Engineering, Salem, Tamil Nadu, India

ABSTRACT

These days, cloud-based storage systems are essential for the processing, sharing, and storing of IoT data. The existing cloud-based design may create serious data leakage or compromise user privacy, not with standing its contribution. The cloud-based architecture, on the other hand, operates under centralised control and heavily relies on a trusted third-party auditor (TPA). The centralised system could fail if there was a single point of failure because the TPA might not be entirely reliable. Thankfully, the decentralised storage paradigm has been more well-known since the development of block chain technology. In addition to eliminating the single point of failure and eliminating the rule of TPA, a decentralised storage system offers various advantages over a centralised control architecture, including low storage costs and high throughput. This research proposes a block chain-based decentralised distributed storage and sharing mechanism with fine-grained access control and end-to-end encryption. Our suggested IoT Chain concept uses the Ethereum block chain as an auditable access control layer and bases fine- grained permission on attribute-based access control (A-BAC) policy. The IoT Chain concept, which combines the Ethereum block chain and the interplanetary file system, is designed specifically for smart contracts (IPFS). For the encryption of shared secret keys between data owners and consumers, we used the advanced encryption standard (AES).

Index Terms - IoT, Blockchain, Hyperledger, IoT Security, IoT with Blockchain.

1. INTRODUCTION

Today, IoT devices are everywhere, in smart homes, wearable devices, smart cities, healthcare, automotive, environment, smart water, and grid applications, etc. IoT solutions are used in many areas for optimizing production and transitioning industries to information technologies. Approximately 46 billion devices will be connected to the Internet of Things bythe end of 2021, according to Juniper Research (Juniper Research, 2021).

Despite the fantastic benefits that IoT technologies provide, assuring the integrity and dependability of the data obtained from these technologies remains a problem that has not yet been fully addressed. It is unfortunately easy to capture and manipulate the data transmitted by many IoT devices. As their capacity for information processing, storage, and networking is constrained, IoT devicesare typically more susceptible to assaults (Jyoti and Amarsinh, 2017)It is deemed necessary usage of different designs, different interfaces, or different environments to ensuresecure communication.

According to Alfonso (2018), it is usually thought to be challenging to ensure that the data produced by IoT devices has not been altered or changed in any manner.where IoT devices send data to shared servers that keep all records centrally.

IoT system security issues may be resolved by blockchain (Dikilitas et al., 2021). In the most basic terms, a blockchain can be defined as a decentralized and distributed ledger technology that contains interconnected records. A record can be added to the blockchain ledger with the approval of the majority of all peers on the network. Blockchain's interconnected blocks nature ensures that the data is protected from tampering. This decentralized structure of the blockchain offers security and privacy (Rui et al. 2019). Blockchain technologies can provide a high level of security, privacy, authentication, and device authorization for the data to be recorded and can be used to secure systems that use IoT devices.

However, there is no clear solution in the literature on how to securely transfer data from IoT devices to a blockchain. The main motivation in our study is that the data transfer between these two environments can be done securely with the integration of IoT and blockchain.

The outline of the study is as follows, summary information about IoT and blockchain is included in section II. In section III, the work done on IoT and blockchain integration is summarized. Section IV discusses the suggested architecture for safely integrating blockchain and IoT, and Section V offers recommendations.

2. BACKGROUND

Internet of Things (IoT)

The density of digital data in our lives has started to increase very rapidly because of continuously online devices. Since these data are actively transferred over the internet, they are always accessible. The technology that enablesthis intensive data transfer between human beings

and devices is called IoT.

Standard IoT devices are heterogeneous devices with embedded sensors that connect over a network. IoT devices are uniquely identified in the network. These devices are generally designed to operate with small memory, limited processing capacity, and low power. Networks act as a bridge between IoT devices and users.

Due to their tiny memory size, low computing power, and low power consumption, Internet of Things (IoT) devices have security flaws in the areas of identity verification, authorisation, and accounting.. However, IoT's seven-tier architecture (Jasmin, 2016) allows participants to develop IoT devices that are compatible with each other by determining layers in which certain types of transactions can be optimized.

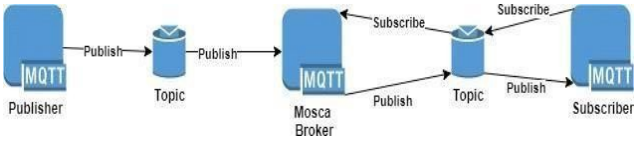


Figure 1. Architecture of MQTT

Data received from IoT devices can be transferred according to protocols (Nitin, 2017). The most widely used one is MQTT. It is an IoT communications protocol that adheres to OASIS standards. It is designed as an extremely light publish-subscribe messagingtransport ideal for connecting remote devices with a small code space and minimal network bandwidth.

The MQTT structure consists of three basic components: "publisher", "subscriber" and "broker". MQTT Publisher posts a message under a specific topic via the MQTT broker. Likewise, the MQTT subscriber can access all the messages published under this topic by subscribing to the topic published by the MQTT publisher via the MQTT broker.

3. BLOCKCHAIN TECHNOLOGY

Blockchain is a peer-to-peer (P2P) network that is decentralised, distributed, unchangeable, and shared to keep track of assets and transactions (Nabil and Claus, 2018). Blocks are among the most crucial ideas in blockchain.. Each block contains an encrypted value containing the block information preceding it.

Transactions are a small task unit usually stored in public records within a block. Records are often only posted to the blockchain with the consent of the vast majority of users who are actively using the network (Gareth and Efstathios, 2015). The data processed in the distributed ledger is recorded in the ledger with the consent of all participants in the network. This is called a consensus mechanism.

The decentralisation structure of Blockchain is the first of its valuable qualities. In a blockchain network, data appears as a distributed ledger database. The same data are kept simultaneously on all other stakeholders in the network. With all these features, blockchain technology provides the highest level of traceability by all stakeholders

Another important contribution of blockchain technology is the transparency it adds to business processes. This increases the trust between stakeholders and ensures accountability. It allows all stakeholders to monitor the blockchain network in real-time.

Data privacy is an important feature of restricted blockchain networks. In restricted blockchain networks, only users authorized by the node's administrator can view data. This ensures that data coming to the blockchain network is protected Integrity issues are substantially eliminated by blockchain networks' public key architecture, which safeguards against data alteration. The participants and the consensus mechanism of a blockchain network are other factors that increase data security.

Smart contracts are pieces of code that define the business processes between stakeholders in blockchain networks. Blockchain technology is used on many digital currencies including widespread Bitcoin and Ethereum cryptocurrencies using smart contracts. In the field of corporate blockchain applications, Hyperledger Fabric network is frequently used.

4. RELATED WORKS

In this section, relevant studies about IoT, blockchain integration, and security will be discussed. At (Nejc, 2019), an approach is presented to integrate IoT and blockchain technologies into supply chain processes. The study proposes a Blockchain- based distributed logistics platform that involves adding actors in a supply chain to the system as a node. A virtual copy of transported property is created with IoT devices on the proposed platform. Other data such as the location, temperature, and humidity of the transported goods are monitored via the virtual copy and recorded on the blockchain.

At (Thomas, 2017), It is explained that the most important benefit in saving data from IoT devices to a blockchain network using smart contracts is that these data can be evaluated and reported to the sender or receiver automatically. The smart contract running in the system corrects the temperature values coming from the sensors and saves them to the blockchain. The REST API is used by mobile clients to communicate with the server. Customers can control the data in the system through these mobile clients.

At (Kristi´an, 2019), a blockchain-based network monitoring and management architecture is proposed. The administrators in the network indirectly control the network devices by recording the changes in the device configuration on the blockchain, where they control the updates in the configurations of the network devices.

At (Seyoung et al. 2017), It is aimed to control and configure IoT devices using blockchain. The proposed system includes a smartphone and three raspberry pi. The three raspberry pi in the system act as a meter, air conditioner, and light bulb, respectively. The user can adjust the policy in the system via the smartphone. Configuration changes made via the smart device are recorded onthe Ethereum network.

At (Mayra et al. 2016), It is pointed out that blockchain's decentralized and change-resistant nature can be used to solve some of theproblems faced by the nature of IoT. The authors present a cloud and fog-based solution to solve this problem. In the study, the IntelEdison Arduino card is used as the IoT device, and the blockchain works separately on the fog and the cloud. In the experiments, the IoT device writes data to the blockchain via a Python server.

Studies in the literature are generally based on IoT and Blockchain integration. However, there is no clarification on how to ensure security in data transfer, which is one of the biggest disadvantages of IoT systems. In section IV, we offer a solution to this issue.

5. PROPOSED APPROACH

This section introduces a straightforward method for securely connecting IoT devices to any blockchain network. Hyperledger Fabric is employed as the blockchain network in our study.

In the proposed architecture presented at figure 2, an IoT device via its sensors captures the data of interest and publish to MQTT message broker. The message broker distributes the received messages to the registered subscribers. In the implementation of our model, "Mosca MQTT broker" is utilised. The subscriber of the data of interest is an application within the blockchain network which is defined as an authenticated user to the blockchain node. This application creates the necessary transaction on the Hyperledger network based on received data.

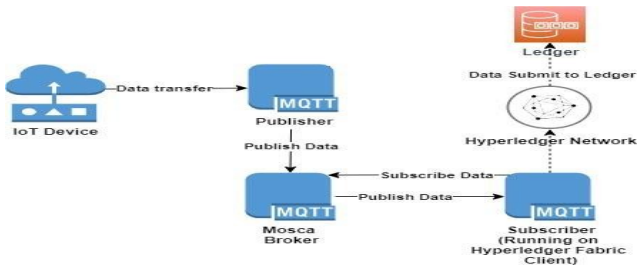


Figure 2. IoT Integration to Hyperledger Network

This approach is simulated in a simple hyperledger network setup. Four Hyperledger Fabric organizations are created on four separate physical machines. Organizations communicate with each other via Docker Swarm Network. Four organizations operate the processes in the network with a channel policy in which they are participants. Three of the four organizations were established to implement the relevant business processes, the fourth organization on the network is Orderer. Organizations in the network each create users to execute the defined processes through smart contracts.

Simulated IOT devices in figure 3 in between organizations trigger events that inform the network of state changes via collected data from sensors. The state changes can be applied to many subjects of interest such as production, transfer of goodsin supply- chain processes as production of a material, transferof materials.

Figure 3. Transaction request flow

With this strategy, it is ensured that the information obtained from the Internet of Things device is delivered to the network across a secure transfer layer via the Hyperledger Fabric user, specified as for the IoT device, preventing unauthorized access and modification of the data. The proposed approach satisfies IoT security problems as discussed below in terms of dimensions of network security.

Confidentiality is ensured with the solutions regarding privacy control and data security offered by the Hyperledger Fabric network (IBM, 2021). Hyperledger Fabric only allows transactions between predefined users via authorized channels which prevents unauthorized users are prevented from making transactions within the network.

The integrity of the data is ensured by keeping data of the network in distributed databases called ledgers. Blockchain ledgers consist of all of the transactions called blocks and that are linked together in a chain. A block that stores a hash value containing the details of the block before it makes this connection. If a peer wants to intervene directly in the Hyperledger Fabric state database, the data cannot be changed, as other peers will not be convinced that this is the case. Having these distributed databases in all stakeholders ensures data integrity.

In terms of Availability, Hyperledger network provides resilience to failures with its distributed nature and the data will be kept continuously accessible for authorized users. In multiple peers carrying the ledger databases, the data will be available over other healthy peers even when one peer is inaccessible. Hyperledger network also provides a high count of transactions per unit time byusing cheaper commodity hardware (Hyperledger, 2021). Identification, Authentication, and Authorization. The MQTT publisher running on the Hyperledger Fabric API server must transact with a previously defined authorized user within the Hyperledger Fabric organizations.

At the same time, additional authentication is provided by using the username and password fields provided by

MQTT for authorization. With this identification provided by the Hyperledger Fabric user identification and MQTT protocol, the data sent through the protocol will be authenticated, and unauthorized access to the data will be prevented. In addition, the optional MQTT TLS support makes it easy to encrypt messages and authenticateclients using authentication protocols.

In terms of Accountability, all elements that will operate in the proposed architecture will be able to perform these operati ons witha specific identity, and these transactions will be accountable by answering questions such as from which IoT device, when and from which Hyperledger Fabric user the transaction was made.

6. CONCLUSION

In conclusion, the proposed framework provides a secure and transparent data sharing mechanism for IoT environments by utilizing cloud computing and blockchain technology. The framework ensures secure data sharing by utilizing blockchain's decentralized and tamper-proof ledger to store data and access control policies. Additionally, the framework providesa reliable and secure data storage solution by utilizing the immutability of the blockchain. The proposed framework can be used in various IoT applications, including healthcare, smart cities, and transportation.

7. REFERENCES

[1] Jasmin G.: Comparison of IoT platform architectures: A field study based on a reference architecture. 2016 Cloudification ofthe Internet of Things (CIoT), pp. 1–6 (2016).

[2] Nitin N.: Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP. 2017 IEEEInternational Systems Engineering Symposium (ISSE), pp. 1–7(2017).

[3] Nabil El I. and Claus P.: A Review of Distributed Ledger Technologies. OTM 2018 Conferences, Vol. 11230, Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 277–288 (2018).

[4] Gareth P. and Efstathios P.: Understanding Modern Banking Ledgers Through Blockchain Technologies. Future ofTransactionProcessing and Smart Contracts on the Internet of Money (2015).

[5] Nejc R.: Distributed logistics platform based on Blockchain and IoT. 52nd CIRP Conference on Manufacturing Systems (CMS), pp. 826–831 (2019).

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Secured E-Voting System Using Two-FactorBiometric Authentication

SARAVANAN O[1], ANUSIYA M[2], KALAIVANI S[3], YOGAVIGNESH V[4]
[1]Professor, [2,3,4]Student
[1,2,3,4]Department of Computer Science and Engineering,
Annapoorna College Engineering, Salem, Tamil Nadu, India

ABSTRACT

In a vastly democratic nation like India, where the people choose the leader, elections are crucial. To overcome this, a biometric validated voting mechanism is introduced in our proposed work. Earlier, slip-based voting mechanisms were available, making fraudulent votes possible. Nowadays, electronic voting machines are also available, making duplicate votes feasible. The suggested system can manage several electronic ballots with different scopes at once, including legislative, municipal, and presidential ones, among others. In terms of both functional and non-functional needs, the system supports election process integrity. The functional requirements built into the proposed system's design mandate well-secured voter identification and authentication procedures through the employment of a combination of straightforward biometrics. Fingerprints and irises are combined and used for authentication in place of a single biometric image, improving security and preventing the casting of fraudulent ballots. By properly including system FLAGs, the system's design ensures that no votes cast in support of a particular candidate are miscounted and lost. To ensure that a voter's vote was cast in favour of the candidate of their choice, voting transparency is maintained throughout the whole electoral process.

Index Terms - Hash Function, Arduino Uno, Biometric, Blockchain, IOT, Machine Learning,and Face Recognition.

1. INTRODUCTION

Elections are the formalised system that helps nations spread democracy[1]. The electoral process is crucial to the nation's future and the future of its citizens. The election process must be carried out in a very well organised manner and should be a trustworthy election because it is a very essential process for everyone concerned. Elections must be conducted in a transparent manner while also taking into consideration voter privacy and security.

The authorities involved in the election process should be fully accountable from the beginning to the end of the process. Forexample, the authorities should not takelonger than necessary to count the votes because doing so increases the risk of vote fraud and the missing of votes, both of which have a negative impact on the nation's elections. Due to several problems like these, sectors dependent on them now needs more trust, which is a major problem.

A centralised authority might have easily manipulated the results of the paper elections. Additionally, the difficulty of timing results announcements hasincreased.

One of the machine learning techniques, face recognition uses a variety of algorithms to recognise faces while also including the study of artificial intelligence.

Face detection is one of the more sophisticated technologies used to locate and measure human faces in digital photos. It exclusively detects facial features and ignores other objects in the image, such as trees, buildings, and bodies.

According to study, the discipline of computer vision is actively exploring human facial perception. In terms of methodology, the initial step is to locate the human face using programmes for picture database management, graphical user interfaces, and video surveillance[18]. If a normalised face image is available, the first step in recognising facial expressions or tracking human faces is to locate them. We incorporate the haar classifier in our projectfor facial recognition.

Blockchain is a straightforward andunchangeable system. It is one method of storing data that afterwards makes it impossible to modify or for any other hacker to hack or game the system. It functions as a ledger of transactions, making a duplicate copy and distributing itover the entire system network.

Due to this important factor, using blockchain to safeguard votes during elections may be an alternative to using traditional methods. It offers one of the clever solutions for central authority issues where each block has data recorded and forms an informational chain.

Since other blocks that have the whole dataare found, it is very difficult to change the information of the data in the blocks. The requirement for using Blockchain enhances information security by keeping all data secure in all blocks; as a result, there is no longer a need for humans to manually preserve and secure the votes[5].

The last node of the chain stores all the information, so we can quickly search for the last node of the chain for the results of the conducted elections. As we are aware, the previous election method takes longer to count the votes

and publish the results tothe public. This has the potential to save a lot of time while also allowing for the delivery of some imperfect outcomes.

One of the more sophisticated sciences and technologies for gathering and evaluating biological data is biometrics. Fingerprint refers to the process of identifying a personvia a finger comparison. It is a well-known biometric that is utilised for computer system authentication. The technology entails taking measurements of and analysing data from human bodily traits like DNA, fingerprints, voice patterns, andhand sizes.

This branch of biometrics was established to broaden the scope of physical identification methods, and it has been put to good use. The biometric approach has been employed by law enforcement as well due to the widespread use of fingerprints and facial recognition as identifiers. Fingerprint scanners and web cams thatswiftly identify people and grant access to everyone were given away by the Election Commission as a result of advances in identifying technology.

2. EXISTING SYSTEM

We currently have paper ballots and electronic voting machines, with paper ballots being an outdated approach that uses paper, pen, and pencil. The votes from an electronic voting machine are stored in theCPU memory.

Then there is fingerprint recognition, which identifies each voter, counts the votes, and guards against fraudulent ballots. The current system is no longer technologically advanced, digital, or secure. The user may select a candidate from the buttons panel to cast their vote if the fingerprint matches their personal information. The results are displayed on LCD for the voters' enjoyment. The existing voting system is vulnerable to numerous risks, including virus attacks, DDOS attacks, vote tampering and manipulation, polling booth capture, andvote altering. Several drawbacks of the current system include:

- Centralized architecture.
- Attack prone.
- Not trustable.
- Non-transparent vote casting process.

3. METHODOLOGY

Proposed System

Election polling could be both an expensive and complicated process. Here, we suggesta brand-new polling method that is cost- effective, secure, and privacy-preserving and uses blockchain technology to store data. Two types of users can use this system: Voter functionality was built into the booth manager and booth manager system where voters go to vote.

Voters must travel to the booth, where the booth manager authenticates them before allowing them to cast their ballots on the voting laptop. Voters are recognised using fingerprint and face recognition authentication, and it is determined whether or not the matching of the fingerprint and face occurs.

Authentication This proposed system offers a means of conduct operations on encrypted data without decrypting it, which, after calculation, can give us results that are comparable to those we would get if we were working directly with raw data.

4. IMPLEMENTATION

The following process occurs:
Information on the user and the booth manager must be given by the election officer. After that, the booth manager sets the voter and fingerprint data. Data will be delivered serially to the Python IDLE oncea fingerprint match has been made.

As soon as the data is received in Python, the face recognition component is combined, the camera is triggered, and the face is matched with the ID. Data will be transferred to the blockchain componentonce both have been matched. voting boothverification. Following casting, each vote is recorded in a block chain database. The election is then over. The output is then shown.

By altering the grey value range and enhancing visual contrast, the process of image normalisation seeks to normalise the intensity levels of pictures. Reduced fluctuation in the grey level throughout the wires is the main goal of normalisation, which also makes subsequent processing operations easier.

Segmentation: Segmentation is necessary to take out the noisy areas and sides of the image. It is based on the calculation of the variance of the grey level. This step enables the extraction phase for biometric data to be optimised and the size of a piece of the image to be minimised.

Face Identification

The Paul Viola and Michael Jones- proposed characteristics are employed by the Haar Cascade, a machine learning (ML)object identification calculation, to identify disagreements in images and videos. This ML-based technology creates a course work from a large number of both positive and negative images. Then, it is applied to identify protests in various images. There are four stages to the calculation:

- Haar Feature Selection
- Making Integral Images
- Adaboost Training

Cascading Classifiers

It is amazing that while being able to recognise practically any object, one can distinguish between faces and body parts inan image. Take face location as an illustration.

To prepare the classifier for computation, a large number of positive photos and negative images without appearance are required. Highlights will be taken out of the

computation. The first thing we do is collect the Haar Features.

A Haar include considers surrounding rectangular districts at a given area in a location window, condenses the pixel controls there, and records the degree of complexity between these sums.

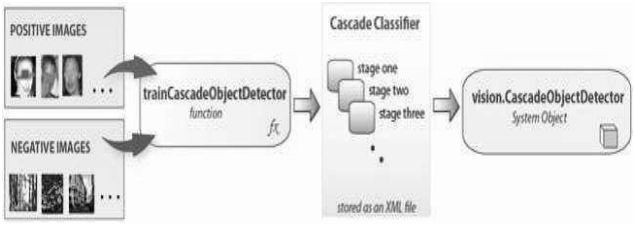


Fig.3: Haar Cascade Classifier

Blockchain:
The final stage involves storing the cast votes as blocks. The hash function that we use is SHA256 (Secure Hashing Algorithm). The value it outputs is 256 bitslong. It is one of the secure hashing algorithms that is frequently used

5. CONCLUSION

In conclusion, our suggested two-factor biometric two-factor protected e-voting system is a safe and effective approach to conduct elections. We make sure that only authorised people are permitted to cast ballots by employing fingerprint and facial recognition technology, and encryption techniques safeguard the integrity and confidentiality of the voting data. The integrity of the election results may be maintained while addressing the security issues with electronic voting with our suggested approach.

6. REFERENCES

- Hanifatunnisa, R., & Rahardjo, B. (2017, October). Blockchain based e-voting recording system design. In *2017 11th International Conference on Telecommunication SystemsServices and Applications (TSSA)* (pp. 1-6). IEEE.
- Krimmer, R., Ehringfeld, A., & Traxl, M. (2010). The use of E-voting in the austrian federation of student's elections 2009. In *4th International Conference on Electronic Voting 2010*. Gesellschaft für Informatik eV.
- "The Geneva Internet Voting System," Internet: https://www.coe.int/t/dgap/goodgovernance/Activities

❑❑❑

Identifying and Forecasting Early Reviewers for Successful Product Marketing on E-Commerce Websites

THANGADURI K[1], JAYAMANI M[2], SARANYA R[3], GOWTHAM J[4]
[1]Associate Professor, [2,3,4]Student
[1,2,3,4]Department of Computer Science and Engineering,
Annapoorna College Engineering, Salem, Tamil Nadu, India

ABSTRACT

Before making a purchasing decision, people are increasingly turning to online reviews as a source of information. Early customer opinions of a product frequently have a big impact on later sales. In this essay, we take the initiative to investigate the traits of early reviewers through the posted reviews on two significant real-world e-commerce platforms, namely Yelp and Amazon. We specifically divide product lifecycle into the early, majority, and laggards stages, which are sequentially ordered. A user that posted a review early is known as an early reviewer. Based on their rating practices, the helpfulness ratings they obtained from other reviewers, and the association between their evaluations and product popularity, we quantitatively describe early reviewers. We have discovered that early reviewers (1) tend to give a higher average rating score and (2) tend to post more enlightening comments. Also, according to our examination of customer evaluations, the popularity of a product may be influenced by early reviewers' ratings and the helpfulness scores they earned. We present a novel margin-based embedding model for early reviewer prediction by considering the review posting process as a multiplayer competition game. Several tests on two separate e- commerce datasets have proven that our suggested method performs better than a variety of industry- standard baselines.

Index Terms - E-Commerce, Herd Behavior, Early Reviewers, Feasibility.

1. INTRODUCTION

The emergence of e-commerce sites enables users to publish or share purchase experiences by posting product reviews, which usually contain useful opinions, reviews, and product feedback. Therefore, most customers will read online reviews before making an informed purchase decision. We call users who post early comments, early comment users. Although early reviewers only contributed a small number of comments, their opinions can determine the success or failure of new products and services. Finding early reviewers is crucial for businesses since their comments can help them modify their marketing plans and enhance their product designs, which will ultimately lead to success. Therefore, in the early promotion stage of the company, early reviewers will become the focus of monitoring and attraction. Marketing professionals have paid close attention to the crucial impact that early comments play in motivating consumers to make purchases.

2. EXISTING SYSTEM

The reality that people are heavily impacted by the decisions of others, which can be described by herd behaviors, has been emphasized in previous studies. The influence of early comments on subsequent purchases can be understood as a special case of the herd effect. When consumers use other people's product evaluations to evaluate the quality of products on the Internet, herding behaviors occurs during online shopping. Unlike existing research on social behaviors, we focus on using large-scale real- world data sets to quantitatively analyses the overall characteristics of early reviewers. On top of that, we formalized the early reviewer prediction task as a competition problem and proposed a sophisticated embedding-based ranking mechanism for it. As far as we know, the tasks predicted by early reviewers themselves have received little attention in the literature

3. PROPOSED SYSTEM

To anticipate early critics, By treating the review publishing process like a multiplayer competitive game, we offer an innovative approach in this work. Only the most aggressive users can join the early reviewers' workforce. Product The competitive process canbe further broken down into pair wise comparisons between two participants. In a two-person contest, the individual with the earlier timestamp will defeat the loser. We suggest employing a margin-based embedding approach, which is motivated by recent advancements in distributed representation learning. The method is to first map users and products to the same embedding space, and then determine the order of a pair of users for a given product according to their respective products. The distance represented by the pr

4. FEASIBILITY STUDY

Analyze the feasibility of the paper at this stage, and propose a business plan, including the overall plan of thepaper and some cost estimates. In the process of system analysis, a feasibility study will be conducted on the proposed system. This is to ensure that the proposed system does not burden the company. In order to conduct a feasibility analysis, it is necessary to understand the main requirements of the system. The

three key factors involved in the feasibility analysis are:

- Economical feasibility
- Technical feasibility
- Social feasibility

Economical Feasibility: This study was conducted to examine the economic impact of the system on the organization. A company's ability to invest money in system development is constrained.. Expenses must be reasonable. Therefore, the developed system is within the budget and can be implemented because most of the technologies used are provided for free. Only need to buy customized products.

Technical Feasibility: This study is conducted to check the technical feasibility, that is, the technical requirements of the system. Any system developed must have no high requirements on the available technical resources. This will lead to high demands on available technical resources. This will lead to high demands on customers. The developed system must have moderate requirements, because the implementation of the system requires only small or no changes.

Social Feasibility: The aspect of the study is to check the user's acceptance of the system. This includes the process of training users to use the system effectively. The user must not feel threatened by the system but must accept the threat from the system. User acceptance only depends on the method of educating users about the system and making it familiar. Their trust must be increased so that he can also put forward some constructive criticisms, which is welcome because he is the end-user of the system.

5. SYSTEM ARCHITECTURE

The following figure depicts the entire system architecture used to characterize and predict early reviewers of product marketing on e-commerce websites.

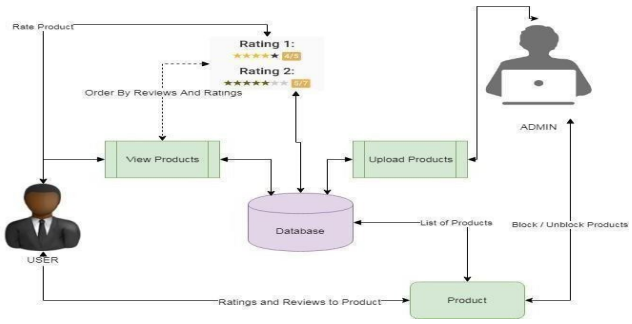


Figure 1: System Architecture

A) Class Diagram

Class diagrams are the main building blocks of any object-oriented solution. It shows the classes in the system, the attributes and operations of each class, and the relationship between each class. In most modeling tools, the class contains three parts. The name is at the top, the attribute is in the middle, and the operation or method is at the bottom. In large systems with manyrelated classes, group classes together to create class diagrams.

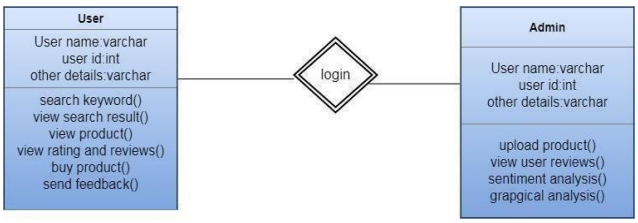


Figure 2: Class Diagram

6. SYSTEM IMPLEMENTATION

There are four modules can be divided here for this paper they are listed as below:

- Upload products
- Product review-based order
- Ratings and reviews
- Data analysis

From the above four modules, paper is implemented. Bag of discriminative words are achieved:

- **Upload Products**: Uploading products is done by the administrator. The authorized person is uploading new arrivals to the system listed by the user. You can upload products and their attributes, such as brand, color and all other warranty details. Uploaded products can be blocked or unblocked by users.
- **Product Review Based Order:** Based on user reviews and ratings of specific products, suggestions for user product views are listed. This project uses the Naive Bayes algorithm to determine whether the sentiment of a given review is positive or negative. Based on the output of the algorithm, suggestions to the user are given. Apply the algorithm and list products on the user side based on positive and negative.
- **Ratings and Reviews:** Ratings and reviews are the main concepts of the paper; the purpose is to find effective product marketing. The main purpose of this paper is to get user reviews based on how users buy or whether to buy. The main findings of the paper are when to give ratings and their effects. This will help users who are willing to buy the same product
- **Data Analysis:** The main part of the paper is to analyses ratings and comments given by users. The product can be analyzed according to the serial number given by the user. The user data analysis of the data can be done through the chart format. The graphics may be different, such as pie charts, bar charts, or some other charts.

7. SYSTEM REQUIREMENTS

This project involves analyzing the design of a few applications to make the applications more user -friendly. For this reason, it is very important to maintain an orderly navigation from one screen to another while reducing the amount of typing that users need to do. To make the

application more accessible, the browser version must be selected to makeit compatible with most browsers.

- **Hardware Requirements:** For developing the application the following are the Hardware Requirements:
  - Processor: Pentium IV or higher
  - RAM: 256 MB
  - Space on Hard Disk: minimum 512MB
- **Software Environment:** For developing the application the following are the Software Requirements:
  - **Python:** Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.
  - **Django:** A Web development framework that saves you time and makes Web development a joy. Using Django, you can build and maintain high-quality Web applications with minimal fuss.
  - **PyCharm:** PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains.It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django.

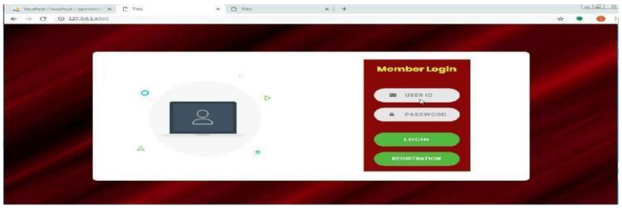


Figure 5: The above figures show about user reviews and ratings

Figure 6: Comparison of various vendors for product

8. CONCLUSION

In conclusion, identifying and forecasting early reviewers is crucial for successful product marketing on e-commerce websites. Early reviewers can provide valuable feedback and influence the purchasing decisions of other customers. By analyzing customer behavior and data, businesses can identify potential early reviewers and incentivize them to leave reviews.

Furthermore, forecasting early reviewers can help businesses plan their marketing strategies and allocate resources effectively. Our methodology can help businesses improve their online presence and increase their chances of success in the competitive world of e-commerce.

9. REFERENCES

[1] J. McAuley and A. Yang, "Addressing complex and subjective product-related queries with customer reviews", WWW, 2016, pp. 625–635.

[2] "Experimental study of inequality and unpredictability in an artificial cultural market," by W.D. J. Salganik M. J. and Dodds P. S., published in ASONAM, 2016, pp. 529–532. "Imagebased recommendations on styles and substitutes," SIGIR (2015), pp.529-532.

[3] "Imagebased recommendations on styles and substitutes," SIGIR (2015), pp. 43–52; J. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel.

[4] "Predicting popularity of twitter accounts through the discovery of linkpropagating early adopters," in CoRR,2015, p. 1512 by D. Imamori and K. Tajima.

[5] "Innovation diffusion and new product growth models: A critical review and research directions," InternationalJournal of Research in Marketing, vol. 27, no. 2, pp. 91-106, by R. Peres, E. Muller, and V. Mahajan

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

An Efficient Secure Data Deduplication and Portability in Distributed Cloud Server Using Whirlpool-HCT and LF-WDO

A R ATHIRA[1], DR. P. SASIKALA[2], DR. R. REKA[3]
[1]Research Scholar, [2]Professor & Head, [3]Associate Professor
[1,3]Department of Computer Science and Engineering, [2]Department of Mathematics,
[1,2]Vinayaka Mission's KirupanandaVariyar Engineering College, Salem, Tamil Nadu, India.
[3]Mahendra College of Engineering, Salem, Tamil Nadu, India

ABSTRACT

Distributed Cloud Computing Storage has come up as a service that can expedite data owners (DO) to store their data remotelyaccording to their application or data or file environment. However, insecure data storage, high uploading bandwidth, integration issues of DCS has breached the trustworthiness of the user to store data. In order to conquer the challenge, the work has developed a data Deduplication and portability-based secure data storage in DCS. The work aids to remove unwanted data and selects the most relevant features to avoid data loss by using GK-QDA Feature Reduction Method and HFG feature selection method. The selected cloud server for the respective data or application is analysed for redundant data by data duplication using a whirlpool hashing algorithm followed by a hash chaining algorithm. Finally, to minimize the integration issues while moving the encrypted data between the DCS, the work has developed an LF-WDO technique. An experimental analysis has showed an enormous result by achieving acomputation time of 2987 ms as compared to the existing methods.

Index Terms - Gaussian Kernel –Quadratic Discriminate Analysis (GK-QDA), hybrid forest genetic algorithm (HFG), Levy Flight – Wind Driven Optimization Algorithm (LF-WDO).

1. INTRODUCTION

As one of the momentous technologies, the distributed storage used in cloud computing has enabled aggregate remote data storage. Hulk and ascendable cloud-based storage is provided for the users from the cloud vendors.[1, 2, 3]. However, the security affairs are still an obstacle for enterprises that caused by the operations on cloud side[4, 5].

Recently many works have been done related to distributed cloud storage (DCS) such as Mass Distributed Storage (MDS) [6, 7] using aFully Homomorphic Encryption (FHE) and ABE as a security policy, Security-Aware Efficient Distributed Storage (SA-EDS) etc,[8, 9]. Even though many techniques has been developed, but still there remains to be a challenge of storing the data securely.

The distributed cloud storage peculiarities results in more liability during data transmissions by malignant interventions or abuse activities[10].After all the risks deriving from different network layers are somewhat fully addressed, therefore, it is a confront obstacle to efficiently secure distributed data in cloud systems. This work has proposed a data Deduplication and portability based protected data storage in distributed cloud computing (DCC)to provide a secure data storage in distributed cloud computing[11].

The carry-over of the paper is trace as follows: Section 2 reviews and discusses the related works based on secure data storage in DCC, Section 3 describes the proposed methodology, the speculative analysis of the proposed methodology is performed in section 4, and finally, Section 5 concludes the proposed method with future scope.

2. LITERATURE SURVEY

Yibin Li et al. [12] developed aSecurity-Aware Efficient Distributed Storage (SA-EDS) model, which was mainly supported by the developed algorithms that included Alternative Data Distribution (AD2) Algorithm, Secure Efficient Data Distributions (SED2) Algorithm and Efficient Data Conflation (EDCon) Algorithm.Even though the approach provided with secure data storage but when the selection of cloud server was inaccurate.

Esther Daniel et al. [13] presented an integrity verification and Deduplication of outsourced data with a lightweight auditing method. The developed scheme combined hashing and symmetric encryption with arenovated distributed hash table data structure, which enabled dynamic operations of the dataefficient, also shortened the communication and computation struggle for integrity verification. The encryption scheme was not that secure to overcome external malicious attacks.

Nabeil Eltayieba et al. [14] developedto provide secure data sharing accompanying the concept of blockchain with attribute-based signcryption in the cloud environment. The strategy satisfied the security obligations such as confidentiality and unforgeability, of cloud computing. Further, by its nature wrong results returned as in the traditional cloud server this smart indenture solved the problem of cloud storage. But the scheme was highly

complex and was inaccurate.

Gagangeet Singh Aujla et al. [15] advanced secure storage, verification, and auditing (SecSVA) of big data in a cloud environment based on Kerberos-based identity verification and authentication, and Merkle hash-tree-based trusted third-party auditing on the cloud. Even though the approach enhanced with data Deduplication but portability issue leaded to slow down of the approach.

3. PROPOSED SECURED DATA STORAGE IN DISTRIBUTED CLOUD COMPUTING

In cloud storage, the data Deduplication method is used to reduce the upload bandwidth and storage zone by removing the data clones from the cloud service provider (CSP) but data Deduplication is a challenge in DCS. To overcome such challenges, we proposed a secure data Deduplication system and portability with distributed cloud server as shown in figure 1.

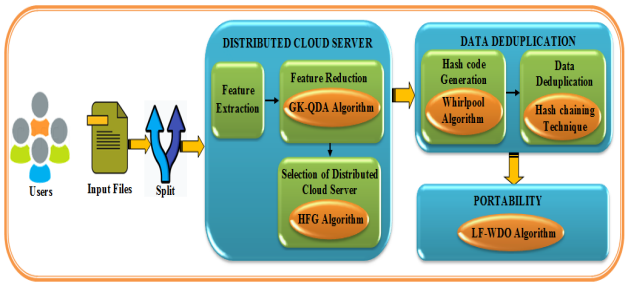


Figure 1: Proposed secure data storage in DCS

Distributed Cloud Server

The distributed cloud server provides storing the data and retrieving it at the time of use. But storing and retrieving is a difficult task in distributed cloud computing due to mass data, irrelevant data, complex data, etc. The work has developed aGaussian Kernel -QDA and Forest Genetic Algorithm for reduction of irrelevant features from extracted features and selection of distributed cloud servers.

Feature Extraction

The initial phase of the proposed work that extricate the various features from file and distributed cloud server. The feature like Task Cost, Speed, Weight, etc. parameters is extracted from the splitted file. In the same way, the features or accessories information like CPU resources, memory resources, and storage resource Processing speed and cycle are extracted from the distributed cloud server.

Feature Reduction

Feature Reduction contributes towards reducing the unwanted features from the extracted features of both file as well as distributed cloud server. As the existing feature reduction technique led to some amount of data loss; to conquer this problem, the work has developed a Gaussian Kernel -QDA Algorithm.

$$\Phi_\kappa(\mathfrak{R}) - log\,\pi_\kappa - \frac{1}{2}\phi_\kappa^T\sum_\kappa\phi_\kappa + \mathfrak{R}^T\sum_\kappa\phi_\kappa - \frac{1}{2}\mathfrak{R}_\kappa^T\sum_\kappa\mathfrak{R} - \frac{1}{2}log|\sum_\kappa| \quad (1)$$

$$\forall(\mathfrak{R}) = \sqrt{\frac{a}{\pi}}.e^{-\omega_i\mathfrak{R}^T} \quad (2)$$

Where ϕ_κ and $\sum_\kappa$ denotes the mean and covariance matrix, $\Phi_\kappa(\mathfrak{R})$ is the given input features, $\forall(\mathfrak{R})$ is the Gaussian kernel function that converts the input features into correlation matrix. The techniques obtain the reduce feature:

$$\mathfrak{F}_i^* = [\mathfrak{R}_1, \mathfrak{R}_2, \mathfrak{R}_3, \mathfrak{R}_4, \ldots\ldots \mathfrak{R}_n] \quad (3)$$

Selection of Distributed Cloud Server

The proposed work is proffered by the feature selection process to tap the cloud server that is by selecting most of the paramount features from the condensed features to reduce the model accuracy, eliminate the insignificant features that may also lead to more computational time. To ameliorate the extracted feature, the work has developed a Hybrid Forest Genetic Algorithm (HFG) stated below:

Step1: Initialize the forest or file features ($\mathfrak{R}_t \in Rand$) with the random trees or cloud server ($\mathfrak{R}_T$) which consists of $(D+1)$ dimensions vector of feature$\mathfrak{F}$. Initially, the age of the trees is set to zero ($\mathfrak{R}_T^{age} = 0$), thereafter, the local solution$\mathfrak{R}_T = LSC(\mathfrak{R}_t)$ is evaluated within the range of $\mathfrak{R}_t \in \Gamma[0,1]$

Step2: For evaluating the local solution crossover $\mathfrak{R}_c \to \lambda_c[\mathfrak{R}_1, \mathfrak{R}_2, \mathfrak{R}_3, \mathfrak{R}_4, \ldots \mathfrak{R}_n]$and mutation is performed on the trees and the mutated value is selected$\lambda_m(r_t) \to [r_1, r_2, r_3, r_4, \ldots r_n]$. Thereafter increment the Tree age by 1.

Step3: Now global seeding is done for the selected tree and random selection of tree is done using GSC parameters$(\mathfrak{R}_T = GSC(r_t))$. Then, the value of each selected variable will be negated (changing from 0 to 1 or vice versa). Based on the trees global search is done over the global space and best tree is updated. Thus, End of the iteration after obtaining a best tree (cloud server) as a substrate of the best selected features (file features)$\mathfrak{F}_s = [r_1^*, r_2^*, r_3^*, r_4^*, r_5^*, \ldots r_n^*]$.

Data Deduplication
Data Deduplication in Cloud server is generated by the hash code for the appropriate split file using Whirlpool Hashing Algorithm. Next, the cloud server checks the hash value availability using the Hash Chaining Technique in the distributed cloud server. If it is available, then the cloud server refers to the stored file location path. If it is unavailable, then the cloud server performs the compression and encryption to store the data. Initially, the selected file features are converted into the hash function using whirlpool given by the file $[r_1, r_2, r_3, r_4, r_5, \cdots r_n]$, the whirlpool hash function is given by:

$$\lambda_m(r_i) \rightarrow [r_1, r_2, r_3, r_4, \cdots r_n]$$

$$H_0' = [r_1^*, r_2^*, r_3^*, r_4^*, r_5^*, \cdots r_n^*] \quad (4)$$

$$H_i' - \Psi(H_{i-1}', r_1^*) \oplus H_{i-1}' \oplus r_1^* - intermediatevalue \quad (5)$$

$$H_s' = f(H_{i-1}', r_{i-1}), 1 \leq i \leq L \quad (6)$$

Where, H_0' is the function that is going o generate hash value H_i' is the hash value generating for the given features, Ψ is the constant function, H_{i-1}', r_1^* denotes the previous hash value chaining with the blocks r_1^* and at last $H_i' = H_s'$ is checked whether the same hash value is obtained or not.

PORTABILITY
Cloud application portability provides an ability to move encrypted data between cloud servers with a minimum level of integration issues with a particular time interval to improve the security of the encrypted data. This phase was done by using Levy Flight – Wind Driven Optimization Algorithm. To replace the selection of random position and velocity, the levy flight technique will be replaced in the WDO algorithm.

Step 1: Initially the feature follows Newton's second law of motion based on temperature and air pressure. Now, based on the moving air force constant, the equation is formulated, such as pressure gradient force $\Omega_{pg} = -\nabla P \delta v$, gravitational force $\Omega_g = r\delta v a$, Coriolis force $\Omega_c = -2\aleph \times u$, friction force $\Omega_f = -r\alpha u$.

Step 2: Based on the equation all the forces are summed together and equated in equation 7 thereafter velocity and position of the air parcel is obtained using equation 8-10:

$$r\frac{u}{\Delta t} = (r\delta v a) + (-\nabla P \delta v) + (-r\alpha u) + (-2\aleph \times u) \quad (7)$$

$$u_{new} = ((1-\alpha)u_{old}) + a(-r_{old}) + \left|\left|\frac{P_{max}}{P_{old}}\right| |(r_{old_{max}}())| + \left[\frac{gu_{old}^{otherdim}}{P_{old}}[]\right]\right| \quad (8)$$

$$r_{new} = r_{old} + (u_{new} \times \hbar_{lf}) \quad (9)$$

$$\hbar_{lf}[r_{new}(k)]r_{old} = \hbar_{lf}[exp(-r_{new}|k|^{\beta})]r_{old} \quad (10)$$

Where, u_{new} denotes the updated air velocity which depends upon the current air velocity u_{old}, α r_{old} is the current search space features of the file with allocated cloud server, P_{max} and P_{old} states the maximum pressure and pressure at the current location ,ΩT and g is the constants, r_{max} states the cloud server facing any problem for storing the file , $\hbar_{lf}$ represents the levy flight distribution for updating the time step ,$[r_{new}(k)]$ is the probabilities of step addition of the random variables and $\beta \in 0,2$.thus from the above technique the files are allocated to new cloud server(r_{new}) due to some issues in the existing servers(r_{old}).

4. RESULTS AND DISCUSSION
The projected framework is validated based on various metrics along with various existing algorithms in order to observe the efficiency of the framework towards secure data storage in DCS.

Performance analysis of the proposed FGA based on various metrics
The analysis elaborates the evaluation based on waiting time, process time, response time, and turnaround time for the proposed FGA along with the existing horse optimization algorithm (HOA), Lion optimization algorithm (LOA), Genetic Algorithm (GA), Forest optimization algorithm (FOA). The evaluation is exhibited in table 1.

Table1: Evaluation of proposed FGA for different metrics

Metrics / Techniques	Waiting Time	Process Time	Response Time	Turn Around Time
HOA	3452	3214	6754	6457

LOA	3124	3020	6186	6124
GA	2865	2651	5214	5647
FOA	2345	2124	4421	5214
Proposed FGA	2143	2005	4002	4876

Based on computation time, the proposedLF-WDO is validated along with the existing methods, such asSpider Monkey Optimization Algorithm (SMOA),Grey Wolf Optimization Algorithm (GWOA),Brownian Motion Bat Optimization Algorithm (BMBAT), andWind Driven Optimization Algorithm (WDO). The graphical analysis of the proposed whirlpool-HCT vs. Hash code generation time is illustrated in figure 3.

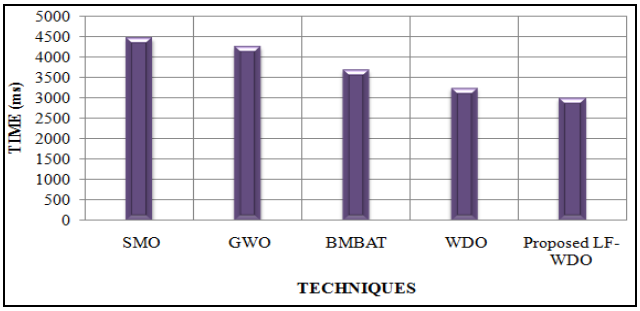


Figure 3: Graphical demonstration of proposed LF-WDO based on computation time

From figure 3, it can be illustrated that the proposed LF-WDO algorithm takes a Computation time of 2987ms to process the entire portability of CS, whereas the existing SMOA, GWOA, BMBAT, and WDO tends to achieve a computation time of 4468ms, 4257ms, 3687ms, and 3241ms, which comparatively consumes more time than the proposed algorithm. Thus, the proposed LF-WDO tends to be robust and secure to solve the integration issues faced between encrypted data and DCS.

5. CONCLUSION

The paper has developed a data Deduplication and portability-based secure data storage in distributed cloud computing. The effort has mainly concentrated on data Deduplication and portability of data in order to reduce the upload bandwidth, storage space, malicious attacks, integration issues, etc. The proposed methods reduce the irrelevant data of file or DCS and select the most appropriate features needed for allocating storage in DCS Using GK-QDA Algorithm and Forest Genetic Algorithm. Thereafter, copied data from CSP is removed using hash code generation and data duplication using the whirlpool algorithm following the hash chaining algorithm. Finally, for avoiding the portability issues, the work has developed an LF-WDO.Experiemental analysis has achieved an better outcome while considering response time of 4002 ms and computation time of 2987 ms as corelated to exisiting expediant methods.

6. REFERENCES

[1] MahdiGhafoorian, DariushAbbasinezhad-Mood, and Hassan Shakeri, "A thorough trust and reputation based RBAC model for secure data storage in the cloud", IEEE Transactions on Parallel and Distributed Systems, vol. 30, no. 4, pp. 778-788, 2018.

[2] MuhammadUsman, Mian Ahmad Jan, and Xiangjian He, "Cryptography-based secure data storage and sharing using HEVC and public clouds", Information Sciences, vol. 387, pp. 90-102, 2017, 10.1016/j.ins.2016.08.059.

[3] Wei Liang, Yongkai Fan, Kuan-Ching Li, Dafang Zhang, and Jean-Luc Gaudiot, "Secure data storage and recovery in industrial blockchain network environments", IEEE Transactions on Industrial Informatics, vol. 16, no. 10, pp. 6543-6552,2020.

[4] Qinlong Huang, Yixian Yang, and MansuoShen, "Secure and efficient data collaboration with hierarchical attribute-based encryption in cloud computing", Future Generation Computer Systems, vol. 72, pp. 239-249, 2017, 10.1016/j.future.2016.09.021.

[5] Yongkai Fan, Xiaodong Lin, Gang Tan, Yuqing Zhang, Wei Dong, and Jing Lei, "One secure data integrity verification scheme for cloud storage", Future Generation Computer Systems, vol. 96, pp. 376-385, 2019, 10.1016/j.future.2019.01.054.

❑❑❑

Sensors in Daily Life

ANJALA MICHAEL[1], SOUMYA GEORGE[2]
[1]Research Scholar, [2]Student
[1,2]Department of Computer Application
St. George's College Aruvithura, Kerala, India

ABSTRACT

Nowadays sensors are very familiar and their need is increasing day by day. People are living with sensors and using them every minute. Our mobile phone is a device that uses a group of sensors together. The use of sensors makes activities easy, faster, and more accurate way. In earlier days we used keypad mobile phones to make phone calls; But now we are using smartphones that are full of sensors. The touch screen in the smartphone is built with a touch sensor. Fingerprint sensors, proximity sensors, light sensors, and so many sensors are used in it. The smartphone is a simple example of a device with different sensors. There are so many devices using sensors for their functioning. Also, sensors are used in different fields for different functionalities. This paper is discussing about different types of sensors and their use, and applications of sensors in different fields.

Index Terms - Sensor, Chemical Sensor, Proximity Sensor.

1. INTRODUCTION

Now, sensors are a common part of our daily life. In the early days, the capacity to comprehend, recognize, value, or respond to something through human effort is referred to as "sense". But now all these actions are done by some physical things called sensors. Humans can take action to the results of these sensors. Sensors are used in different fields for different purposes. The use of sensors causes to decrease in the workload. It makes the results more accurate and quicker. It is sensitive to every moment and action, and it makes immediate responses to those actions in the form of alarms, disconnection of electricity, etc. There are various types of sensors available for different purposes. The following sections are discussing about different types of sensors and their applicationsin different fields. Figure 1 shows the block diagram of a simple sensor.

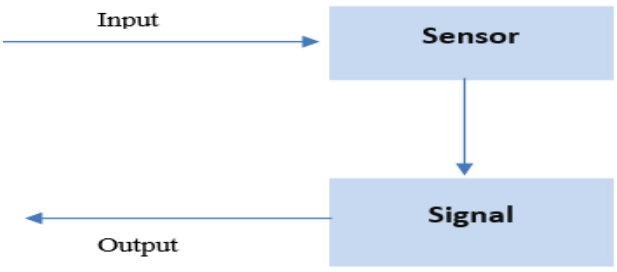


Figure 1: Block Diagram of Sensor

2. SENSORS

Sensoris a device that is used to sense a signal and convert it into information. It recognises and reacts to certain kinds of physical environment inputs. The human body is the best example. Sensory organs in the body act as sensors and sensitive information in the form of impulses to the brain. Nowadays, we are using different types of sensors in different fields. That makes our work easy to handle [1].

2.1 TYPES OF SENSORS

There are different types of sensors for various purpose.

Table 1: Types of Sensors

Sl. No	Sensor	Senses	Using Areas
1.	Temperature sensor	Temperature	Computers, mobile phones, automobiles, air conditioning systems, industries, etc.
2.	Proximity sensor	Presence of an object	Mobile Phones, Cars (Parking Sensors), industries (object alignment), Ground Proximity in Aircraft, etc.
3.	Infra-Red sensor	Proximity and object detection	Mobile Phones, Robots, Industrial assembly, automobiles, etc.
4.	Ultrasonic sensor	Measures the distance of a target object by sound waves	Robotic sensing, Loop control, Stacking height control, Liquid level control, Presence detection
5.	Light / Photo sensor	Measures the amount of light received	Mostly used in cell phones and tablets
6.	Smoke and gas sensor	Smoke and gas	Offices and industries
7.	Alcohol sensor	Alcohol	Breathalyzer device
8.	Touch sensor	Touch of a	All touch screens,

		finger or a stylus	mobile Phones, Tablets, Laptops, etc.) and trackpads of laptops
9.	Colour sensor	Detects colors	Image processing, color identification, industrial object tracking, etc.
10.	Humidity sensor	Measures relative humidity and temperature	Weather forecasting systems
11.	Tilt sensor	Detect inclination or orientation	Used to measure angles or slopes of an object

2.2 APPLICATIONS OF SENSORS

Sensors are used in different fields. Each field has many specific applications. Some of the fields where sensors are used are listed below.

- Agriculture
- Medical care
- Fire and rescue

Agriculture: There are different kinds of sensors used in agriculture, and most of them are used for smart farming. By adjusting to changes in the environment, these sensors data, help farmers to track and improve their crops[2]. Weather stations, drones, and agricultural robots all have these sensors attached. Mobile applications created especially for the purpose can be used to control them. With the aid of a mobile phone software, they can be managed wirelessly using either Wi-Fi directly or cellular towers and cellular frequencies. [3]. Some sensors that are used for smart farming are:

- Location Sensors
- Optical Sensors
- Electro-chemical sensors
- Mechanical Sensors
- Dielectric soil moisture sensors
- Airflow sensors

Location sensor- Location sensor determines the latitude, longitude, and altitude of agricultural fields. This is accomplished with the aid of GPS devices. [3] [2].

Optical sensor- Optical sensors are used to measure the properties of soil. Figure 2 shows the optical sensor. It uses light to measure soil properties. They are attached to satellites, drones, orrobots to determineclay in the soil, organic matter, and water content[3][2].

Figure 2: Optical Sensor[4]

Electro-chemical Sensors: By identifying particular ions in the soil, electrochemical sensors assist in gathering chemical information about the soil. They are giving details about the soil's pH and nutrient content[5]. The following figure 3 shows an electrochemical sensor that is used in agriculture.

Figure 3: Electro-chemical Sensors[6]

Mechanical Sensors -Mechanical sensors are used to measures oil compaction and mechanical resistance[7][2].

Dielectric Soil Moisture Sensors: Dieletric sensors measure the level of moisture content by measuring the dielectric constant of soil[2]. It can have two electrodes as shown in figure 4 which measures the moisture content.

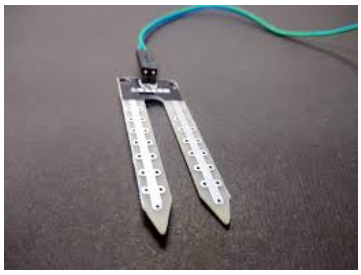


Figure 4: Dielectric Soil Moisture Sensors[8]

Airflow Sensors -Airflow sensors are used to measure airflow. They are used in fixed position or mobile mode[9][2]. Figure 5 shows an airflow sensor.

Figure 5: Airflow sensor [10]

Medical care

In the medical field, sensors are used to erase the difficulty to take medicines. Sensors are used as medicine dispensers to supply the required medication at the right time and serve as a signal to remind people to take their medications [11].There are various sensors used in healthcare applications are:

- Biosensors
- Chemical sensors
- Flow sensors
- Fingerprint sensors
- Force sensors

- Heart rate sensor/ pulse rate sensors
- Humidity sensors
- Hour monitor sensor
- IR sensors
- Image sensors
- Level sensors

Biosensor-The clinical therapy, pharmacy, biomedical, and healthcare industries all use biosensors. Biosensors are effectively used for managing human health, identifying diseases, preventing them, rehabilitating patients, and monitoring their health. Microorganisms such as bacteria, viruses, and parasites can also be found using biosensors. It is frequently utilised in clinical labs today. [12].

Chemical sensor-Chemical sensors are commonly used sensors with uses in the healthcare industry. Due to their high performance, portability, simplicity, and low cost, electrochemical sensors are starting to be used in a variety of analytical, medical diagnostic, and screening applications. [13]. The following figure 6 is used in healthcare devices.

Figure 6: Chemical Sensor [10]

Flow sensors-The typical equipment for detecting airflow is a flow sensor. It measures and controls the flow rate and exact dosing of liquids and gas. It is commonly used in ventilators [14].

Fingerprint sensors- Fingerprint sensors are used to scan fingerprints. Figure 7 is a fingerprint scanner that is used to scan the fingerprint.It is used in medicine to quickly and readily ascertain the patient's previous medical history. Medical records are very confidential for each patient so this fingerprint recognition technology is secured by many cryptographic algorithms [15].

Figure 7: Fingerprint scanner[16]

Force sensors-Compression, force, strain, and load are all measured using force sensors. Fluid monitoring applications, dialysis tools, endoscopic surgery, physical rehabilitation equipment, orthopedics, ECG monitors, and MRI, USG, CT, and PET scanning devices are all examples of uses for this in the medical sector. [17].

Fire and rescue

Sensors are used in many fields. They are also used to detect the presence of fire, smoke, and gas. In the past few days, buildings are built with fire extinguishers and buckets to turn off the fire. But nowadays buildings are made with a good fire-controlling system, more than a fire extinguisher and buckets Fire Alarm Systems, Smoke detectors, Heat detectors, Fire hydrant Systems, and Fire suppression systems are used. Most of them are working using sensors. Here is some equipment used for this purpose.

- **Smoke detector**- The smoke or smoke alarm has two major parts, a sensor, and an alarm. The sensor detects or senses the presence of smoke or fire, it produces an alert alarm to move back from that area and take the necessary actions to turn off the fire [18]. The following figure 7 is a smoke detector that is installed in buildings.

Figure 7: Smoke detector [19]

- **Heat alarm**- The heat alarm is a device that is a special safety device that measures the temperature rise. It measures the temperature rise and gets activated/alert alarm activates when the heat is turning to fire. Heat alarm does not respond to smoke and it is specially designed to protect property and devices. Figure 8 is a heat alarm that is installed in modern buildings and hospitals.

Figure 8: Heat Alarm [20]

- **Fire detector**-A fire detector is also known as a carbon monoxide (CO) detector. These electronic detectors are sensing the outbreak of fire through the level of carbon monoxide in the air [21].

3. CONCLUSION

Sensors are part of our daily life. People are living in the world of sensors. Employees are daily using punching

machines to mark their presence during duty time. The fingerprint sensor is used in it to perform its functions. This study looking at various sensor types and their uses in various areas. There are many more fields where sensors are used, other than the fields that are mentioned in this paper. Sensors are used for the protection of common people from attack by wild animals, and the protection of wild animals from climate change, deforestation, and wild monitoring. The use of sensors gives a perfect result that decreases physical efforts and increases the accuracy of its function.

4. REFERENCES

[1] P. G. R. S. G. Dr.Bhagwati Charan Pater, Advances in Modern Sensors, IOP Publisher, 2020.

[2] A. P. Vivek Parashar, "Study of Various Sensors sed in Farming," *Engineering and Technology Journal for Research and Innovation,* vol. 2, no. 2, pp. 43-47.

[3] Y. C. S. S. Pengxiang Hu, "Low-cost spectrometer accessory for cell phone based optical sensor," in *IEEE*, 2015.

[4] "India Mart," [Online]. Available: https://www.indiamart.com/proddetail/optical-sensors-4429489888.html. [Accessed 21 November 2022].

[5] Y. Y. L. K. S. H. C. L. H. M. N. S. Tan Kong Yew, "An electrochemical sensor ASIC for agriculture applications," in *IEEE*, 2014.

[6] "MOUSER," [Online]. Available: https://www.mouser.in/new/amphenol/amphenol-sgx-sensortech-sgx-4x-sensors/. [Accessed 21 November 2022].

[7] Z. Q. Z. Z. P. S. L. Sun Yurui, "Measuring Soil Physical Properties by Sensor Fusion Technique," in *IEEE*, 2007.

[8] "AUTO DESK," [Online]. Available: https://www.instructables.com/Arduino-Soil-Moisture-Sensor/. [Accessed 21 November 2022].

[9] S. Schriber, "Smart Agriculture Sensors: Helping Small Farmers and Positively Impacting Global Issues, Too," [Online]. [Accessed September 2022].

[10] "MIT tech review," [Online]. Available: https://www.technologyreview.com/2009/06/23/212122/cheaper-chemical-sensor-

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Pollution Control System & Public Safety Protection Using IoT and Big Data Privacy

DR. G. KARTHIK[1], MRS. T.GEETHA[2], C. SIVAKUMAR[3]
[1]Professor, [2]Assistant Professor
[1]Department of Information Technology, [2]Department of Compute Science & Engineering
[1]Karpagam College of Engineering, Coimbatore, Tamil Nadu, India
[2,3]Vinayaka Mission's Kirupananda Variyar Engineering College, Salem, Tamil Nadu, India

ABSTRACT

To control the pollution and manage public safety protection from pollution through IoT and Big data. The Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Humankind, moving to a period centered upon improvement has overlooked the significance of supportability and has been the real guilty party behind the rising pollution levels in the world's air among all other living life forms. The pollution categorized into air pollution, water pollution and noise pollution. The pollution levels at certain spots have come to such high degrees that they have begun hurting our very own. An IoT based pollution observing framework incorporates a sensor interfaced to a outfitted with an WLAN connector to send the sensor perusing to a Thing Speak cloud. Further extent of this work incorporates an appropriate AI model to foresee the pollution level and an anticipating model, which is fundamentally a subset of prescient displaying. As age of poisonous gases from ventures, vehicles and different sources is immensely expanding step by step, it winds up hard to control the dangerous gases from dirtying the unadulterated air, water related pollution & land related pollution. In this system a practical pollution observing framework is proposed. This framework can be utilized for observing pollutions in conduct of specific territory and to discover the pollution peculiarity or property examination.

The obligated framework will concentrate on the checking of pollution poisons concentrate with the assistance of mix of Internet of things with wireless sensor systems. The investigation of pollution quality should be possible by figuring Air Quality Index (AQI), Water Quality Index (WQI) and Land Quality Index (LQI). This system attempts to save the natural resources available for public safety protection kind by continuously checking the quality air, monitoring the status of the soil, the pollution can be controlled and thereby increase the public safety. Also, by knowing the air, water, land moisture and temperature volume of contents are maintained through big data.

Index Terms -.***

1. INTRODUCTION

Pollution can be characterized as nearness of moment particulars that bothers the working of common procedures and furthermore creates unfortunate wellbeing impacts. In another way contamination can influence the characteristic periodicity and furthermore can irritate the wellbeing of person. As modernization and automation is becoming in all respects widely Pollution is likewise getting presented everywhere way.

It has been seen that in mechanically creating or created nations human wellbeing get significantly influenced due to Air Pollution and Water pollution where there is no framework to screen it or monitor it .

In late explores it has been demonstrated that there is a high connection batten's climatic toxins and maladies like asthma and lung related ailments. Air Pollution and Water pollution is currently a noteworthy worry over the globe and WHO has built up specific rules to confine the cutoff points of specific gases like O3, NO2, SO2.

The Air Quality Index and Water Quality Index estimation and Pollution observing are mostly done surface stations that are essentially exact and precise. They

2. EXISTING SYSTEM

The Existing system has know internet are inter-connected devices to the internet. To fulfill the need of flourishing monitoring system, in our project is establishing a network called Internet of Things, in which sensing devices are connected.

Air pollution is not only natural medical matters impact on creating nations alike. The strong effect of air pollution on wellbeing are extremely mind blowing as there are a broad area of sources and their particular influence differ from one another. The synthetic substances reason an assortment of mankind and natural medical issues enlarge

in air contamination impacts on condition also on human wellbeing. To screen this contamination WSN framework is expressed and SO2), and packs them in a casing with the GPS physic distribution, time, and date. The reason is to send the Pollution-Server by means of zig bee device. The pivotal-Server is interact to Google Maps to show the area of equipment. It can associate database server to the Pollution-Server for putting away the toxins range for future utilization by different user , for example, condition security offices vehicles.

DRAWBACKS OF EXISTING SYSTEM

- No devices are connected
- Maintenance expenses are high
- Not Expertise
- No Public Safety
- Slow process
- Highly watchable
- No proper measurement reading for air and water

3. PROPOSED SYSTEM

The proposed system focuses IoT for the most part manages associating shrewd gadgets (implanted hardware gadgets) to Itb by tackling the upside of OSI layered Architecture.

CLIENT CONTROL MODULE

Client control module is intended to cleanup and "hide" options from clients that could potentially break a site after site handoff. The first iteration/release of this module simply provides a settings form to select certain themes which should be hidden from view on the theme listing page. This module is inter connected with the IoT and server control module.

SERVER CONTROL MODULE

Server control module can control the server upto three positions with push button switches or toggle switches. This can be used for a variety of different applications such as:

- Opening level crossing barriers
- Opening goods shed doors
- Opening gates
- Switching points

AIR MEASUREMENT MODULE

This module is used to air measurement module comprises an antenna, adapted to receive a first measuring signal from a device under test or adapted to transmit a second measuring signal to the device under test. The readings based on the Air Quality Index measuring values.

WATER MEASUREMENT MODULE

This module is used to water level sensor module. This water level sensor module has a series of parallel exposed traces to measure droplets/water volume in order to determine the water level. Very Easy to monitor water level as the output to an analog signal is directly proportional to the water level. The readings based on the Water Quality Index measuring values.

LAND /SOIL MEASUREMENT MODULE

This module is used to measure the volumetric water content in soil. Since the direct gravimetric measurement of free soil moisture requires removing, drying, and weighing of a sample, soil moisture sensors measure the volumetric water content indirectly by using some other property of the soil, such as electrical resistance, dielectric constant, or interaction with neutrons, as a proxy for the moisture content. The readings based on the land or soil measuring values.

IoT'S CONNECTED MODULE

This module is used to connection of IoT allows objects to be sensed and controlled remotely across existing network infrastructure, creating opportunities for more direct integration between the physical world and computer-based systems, and resulting in improved efficiency, accuracy and economic benefit. This module is used to interconnection of the client and server module.

PUBLIC SAFETY PROTECTION MODULE

This module is used to create the safety to public from the dangerous. The main aim is to protection of public and give the guidelines to handle the hazards situation.

PROGRESS REPORT CONTROL MODULE

This module is used to create the modules and sub module reports. The progress report is to control the time series.

4. IMPLEMENTATION

Implementation is the stage, which is crucial in the life cycle of the new system designed. **"A CLOUD ASSISTED ZIGBEE-BASED ZOO-ANIMAL HEALTH MONITORING SYSTEM USING BIG DATA AND IOT SERVICES"** is used to implement the experimental validation of animal health monitoring system (AHMS) which is capable to the measuring of body temperature, rumination, and heart rate parameters with environmental parameters (surrounding temperature and humidity). The system is based on the IEEE 1451.2, IEEE 802.15.4, and IEEE1451.1 standards. The PIC18F4550 microcontroller and XBee-PRO S2 module were used to the development of AHM system. The four sensor module such as body temperature, heart rate, surrounding humidity and temperature and rumination has been successfully developed.

They measuring parameters will be helpful to analyze the animal disease or health condition of the animal. The proposed system designed front panel. The front panel of the AHM system handles functions of the measuring parameters such as settings the time interval, start but-ton (ON), OFF, data saves for the access memory of the PC or in the data base, and a digital and a graphical output. Here the developed GUI module can perform for four sensing module and display the seven valuable physiological and behavioural parameters. The USB slot of the PC is present the 100mA at 5V and it does not require any external

power source in the sink module during the experiments. The power consumptions in the AHM system is depend only on the wireless sensing modules. During experiment, the 11.1V battery (rated 350mAh) is used. The each sensor module could be run incessantly for 60 hours without necessitate recharge.

5. CONCLUSION

The Arduino as a motherboard is chosen as the processor. Temperature sensors, smoke sensors, and radiation sensors are the sensors connected to the processor. The smoke sensor senses the smoke if it is greater than 613 it informs the processor and the processor immediately sends a message through GSM with the temperature recorded at that time. The same process will be done if the radiation sensor senses the radiation above the range 250.

When the pollution exceeds the threshold value a message is sent through GSM to the authorized person. Immediately after receiving the message from GSM the power supply of the industry can be cut off through IOT.

The IOT based pollution monitoring and controlling using the Arduino motherboard is connected to the system is designed to sense the smoke, temperature, and radiation. If any pollution is detected then the power supply of that industry will cut off. This will prevent further emissions of pollution. This is a robust system which is very useful industries because of the increasing pollution due to increase in industries. The results of this project are accurate and hence can be implemented in any industries for the safety of workers and the environment. Each and every industry whether small scale or large scale should and must have this system to monitor the emissions.

This system includes sensors that detect the parameters causing pollution. Whenever there is an increase in the level of these parameters the sensors sense the situation it sends as a message to authorized person through GSM. The authority can cut off the power supply of the polluting industry through IOT.

6. REFERENCES

[1] Pollution Control Technologies – Vol. I - Pollution Control Technologies - B. Nath and G. St. Cholakov ©Encyclopedia of Life Support Systems (EOLSS) review on engines, fuels, illustrated with numerous case studies].

[2] Glassman, I. (1996), Combustion, 3rd ed., Academic Press, Inc., London, UK. [Thermodynamic and chemical fundamentals of combustion].

[3] Handbook of Air Pollution from Internal Combustion Engines (1998), Ed. E. Sher, 663 pp. San Diego, CA, USA: Academic Press. Heck R. M. and Farrauto R. J. (2002).

[4] Catalytic Air Pollution Control: Commercial Technology. 416 pp. New York, USA: John Wiley & Sons [Catalysts for pollution control].

[5] http://www.epa.gov/ and http://eea.eu.int/ [the websites of the US and the European agencies with abundance of information and links]

[6] http://www.epa.gov/oeca/sector/ [profiles and reviews of the industries covered in the Theme] http://www.wikipedia.org/wiki/ [A comprehensive encyclopedia explaining numerous themes of physics, chemistry and other sciences]

[7] http://www.wiley-vch.de/vch/software/ullmann/. Ullmann's Encyclopedia of Industrial Chemistry – 7th edition.

[8] Kalhammer, F. R., Prokopius, P. R., Roan, V. P.. and Voecks, G. E. (1998). Status and Prospects of Fuel Cells as Automobile Engines. A Report of the Fuel Cell Technical Advisory Panel, Section II, 19 pp., Prepared For State of California Air Resources Board, Sacramento, California, USA. [An expert report on fuel cells vehicle technology.].

[9] Niessen, W. R. (2002). Combustion and Incineration Processes, 3rd edition, New York: Marcel Dekker; [Basic reference covers the technology of waste incineration systems from a process viewpoint, with attention to the chemical and physical processes.]

[10] Perry, R. H.; Green, D. W. eds. (1997). Perry's Chemical Engineers' Handbook, 7th ed., New York, NY, USA: McGraw-Hill. Poling, B. E.; Prauznitz, J. M.; O'Connel, J. P. (2001). Properties of Gases and Liquids, 5th ed., New York: McGraw-Hill, Inc.. [An extensive amount of experimental data with proven estimation techniques and generalized correlations.]

[11] Wakefield, E. H. (1998) History of the Electric Automobile: Hybrid Electric Vehicles. 332 pp., Society of Automotive Engineers (SAE), Warrendale, Pa. USA. [Developments in the electric automobile conce

❑❑❑

Overload Avoidance for Vigorous Virtual Contraption Resource Allocation Using Green Computing

GNANAVEL. N
Department of Compute Science & Engineering
Vinayaka Mission's Kirupananda Variyar Engineering College, Salem, Tamil Nadu, India

ABSTRACT

Green Computing develops a resource allocation system that can avoid overload in the system effectively. A resource provisioning mechanism is required to supply cloud consumers a set of computing resources for processing the jobs and storing the data. Cloud providers can offer cloud consumers two resource provisioning plans, namely short-term on-demand and long-term reservation plans. Pricing in on-demand plan is charged by pay-per-use basis. The under-provisioning problem can occur and it can be solvedby provisioning more resources with on-demand plan to fit the extra demand. It is important for the cloud consumer to minimize the total cost of resource provisioning by reducing the on-demand cost and oversubscribed cost of under provisioning and overprovisioning. To overcome, ExtendingPhase scheme is used which shares the reservedunused memory to other customers or this unused memory are allocated to the same person to access the other open-source application in same Resource Provisioning application. The Service Level Agreement (SLA) based super-scheduling approach promotes cooperative resource sharing and minimize total cost of resource provisioning in a certain period of time. Using SLA, users can share memory to any other users through pay per basic or open source and user can use unused memories to use other applications. The number of users can be served simultaneously. The proposed mathematical analysis will be useful to the consumers for the management of virtual machines in computing environment.

Index Terms - Green Computing, Virtual Machines.

1. INTRODUCTION

The elasticity and the lack of upfront capital investment offered by cloud computing is appealing to many businesses. There is a lot of discussion on the benefits and costs of the cloud model and on how to move legacy applications onto the cloudplatform. Here we study a different problem: how can a cloud service provider best multiplex its virtual resources onto the physical hardware? This is important because much of the touted gains in the cloud model come from such multiplexing. Studies have found that servers in many existing data centers are often severely underutilized due to overprovisioning for the peak demand. The cloud model is expected to make such practice unnecessary by offering automatic scale up and down in response to load variation. Besides reducing the hardware cost, it also saves on electricity which contributes to a significant portion of the operational expenses in large data centers.

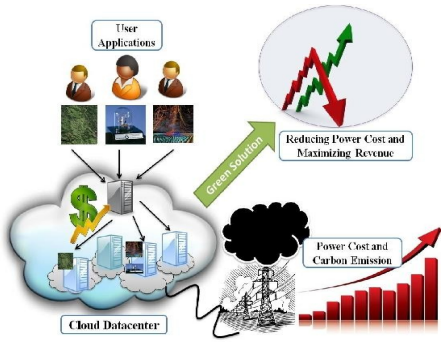


Fig. 1: Cloud Datacenter

Virtual machine monitors (VMMs) like Xen provide a mechanism for mapping virtual machines (VMs) to physical resources. This mapping is largely hidden from the cloud users. Users with the Amazon EC2 service, for example, do not knowwhere their VM instances run. It is up to the cloud provider to make sure the underlying physical machines (PMs) have sufficient resources to meet their needs. VM live migration technology makes it possible to change the mapping between VMs and PMs While applications are running. The capacity of PMs can also be heterogeneous because multiple generations of hardware coexist in a data center. A policy issue remains as how to decide the mapping adaptively so that the resource demands of VMs are met while the number of PMs used is minimized.

When the resource needs of VM's are heterogeneous due to the diverse set of applications they run and vary with time as the workloads growand shrink.

2. GREEN COMPUTING

The design and implementation of an automated resource management system thatachieves a good balance between the two goals. Two goals are Overload Avoidance and Green Computing.

- **Overload Avoidance:** The capacity of physical machines should be sufficient to satisfy the resource needs of all VMs running on it. Otherwise, the PM is overloaded and can lead to degraded performance of its VMs.

- **Green Computing:** The number of physical machines used should be minimized as long as they can still satisfy the needs of allVMs. Idle PMs can be turned off to save energy.

A resource allocation system that can avoid overload in the system effectively while minimizing the number of servers used. We introduce the concept of "skewness" to measure the uneven utilization of a server. By minimizing skewness, we can improve the overall utilization of servers in the face of multidimensional resource constraints. An important issue when operating a load-balanced service is how to handle information that must be kept across the multiple requests in a user's session. If this information is stored locally on one backend server, then subsequent requests going to different backend servers would not be able to find it. This might be cached information that can be recomputed, in which case load-balancing a request to a different backend server just introduces a performance issue.

A variety of scheduling algorithms are used by load balancers to determine which backend server to send a request to. Simple algorithms include random choice or round robin. More sophisticated load balancers may take into account additional factors, such as a server's reported load, recent response times, up/down status (determined by amonitoring poll of some kind), number of active connections, geographic location, capabilities, or how much traffic it has recently been assigned.

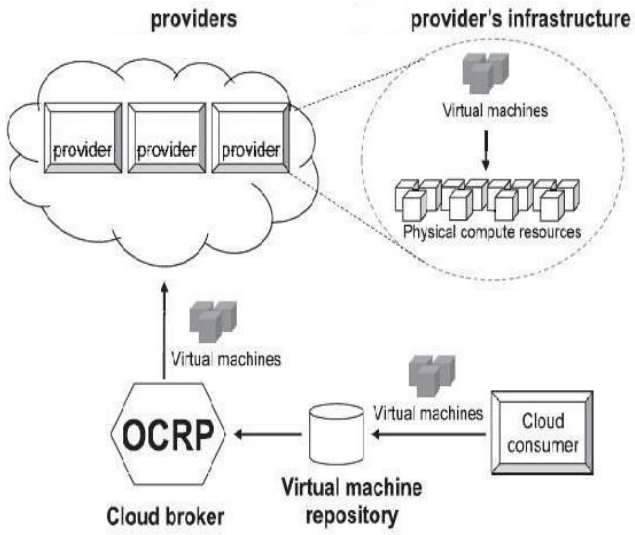


Fig. 2: Virtual Machine Migration

3. TECHNIQUES SKEWNESS ALGORITHM

Application delivery controllers have evolved from basic server load balancing functional units to fully integrate with cloud workflows and provisioning systems so that they help users enable fast roll-out of new applications to a mobilized work force, improve end-user satisfaction, and reduce the time and cost of application deployment We introduce the concept of skewness to quantify the unevenness in the utilization of multiple resources on a server. Skewness is a statistic that is used to measure the symmetry of the distribution for a set of data. The skewness of an analysis domain iscalculated as follows:

Skewness = K_3 / E_{SD}^3

where:

$K_3 = [n \cdot S_i (E_i - E_n)^3] / [(n-1)\cdot(n-2)]$ if n 3;
$K_3 = 0$ if $n < 3$;

E_{SD} Standard_ deviation (Standard Deviation of the Energy)

$$E_{SD} = \sqrt{\sum_i \frac{(E_i - E_n)^2}{n-1}}$$

summation is over all samples i in the region.

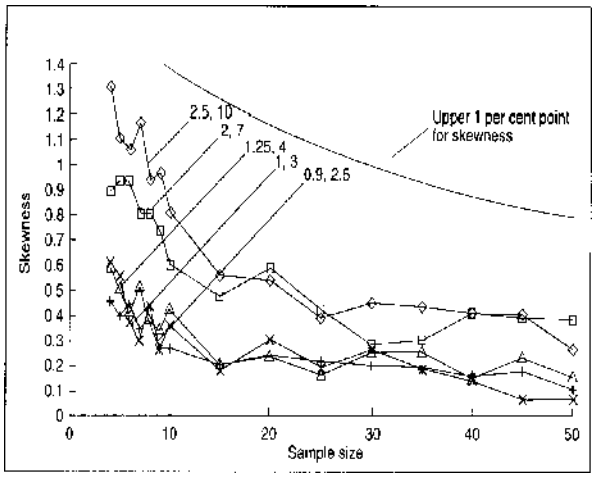


Fig. 3: Skewness

4. VIRTUAL MACHINE MIGRATIONS

VM that can reduce the server's Usage the most. In case of ties, we select the VM whose removal can reduce the skewness of the server the most. For each VM in the list, we see if we can find a destination server to accommodate it. The server must not become a hot spot after accepting this VM. Among all such servers, we select one whose skewness can be reduced the most by accepting this VM. Note that this reduction can be negative which means we select the server whose skewness increases the least. If a destination server is found, we record the migration of the VM to that server and update the predicted load of related servers. Otherwise, we move onto the next VM in the list and try to find a destination server for it. As long as we can find a destination server for any of its VMs, we consider this run of the algorithm a success and then move onto the next hot spot. Note that each runof the algorithm migrates away at most one VM from the overloaded server. This does not necessarily eliminate the hot spot, but at least reduces its temperature. If it remains a hot spot in the next decision run, the algorithm will repeat this

process. It is possible to design the algorithm so thatit can migrate away multiple VMs during each run. But this can add more load on the related servers during a period when they are already overloaded.

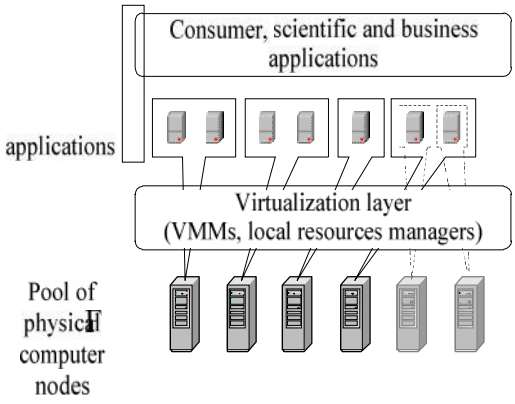
ON-DEMAND PLAN

Pricing in on-demand plan is charged by pay-per-use basis (e.g., 1 day). Therefore, purchasing this on-demand plan, the consumers can dynamically provision resources at the moment when the resources are needed to fit the fluctuated and unpredictable demands.

RESERVATION PLAN

Pricing is charged by a onetime fee (e.g., 1 year) typically before the computing resource will be utilized by consumer. With the reservation plan, the price to utilize resources is cheaper than that of the on-demand plan. In this way, the consumer can reduce the cost of computing resource provisioning by using the reservation plan

EXTENDING PHASE



The over provisioning problem can occur if the reserved resources are more than the actual demand in which part of a resource pool will be underutilized. It is important for the consumer to minimize the total cost of resource provisioning by reducing the oversubscribed cost, to over the above problem we provide new Extending Phase scheme in this scheme the reserved unused memory can shared by other customers or this unused memory's are allocated to the same person to access the other open source application in same Resource Provisioning application.

5. CONCLUSION

Clouds are essentially Data Centres hosting application services offered on a subscription basis. However, they consume high energy to maintain their operations.As energy costs are increasing while availability dwindles, there is a need to shift focus from optimising data centre resource management for pure performance alone to optimising for energy efficiency while maintaining high service level performance.It is important for the cloud consumer to minimize the total cost of resource provisioning by reducing the on-demand cost and oversubscribed cost of under provisioning and overprovisioning. To achieve this goal, the optimal computing resource management is the critical issue.We propose Green Cloud computing model that achieves not only efficient processing and utilisation of computing infrastructure, but also minimise energy consumption.

6. REFERENCES

[1] M. Nelson, B.-H. Lim, and G. Hutchins**, "Fast Transparent Migration for Virtual Machines,"** Proc. USENIX Ann. Technical Conf., 2005.

[2] M. McNett, D. Gupta, A. Vahdat, and G.M. Voelker, **"Usher: An Extensible Framework for Managing Clusters of Virtual Machines,"** Proc. Large Installation System Administration Conf**.** (LISA '07), Nov. 2007.

[3] T. Wood, P. Shenoy, A. Venkataramani, and M. Yousif, **"Black-Box and Gray-Box Strategies for Virtual Machine Migration,"** Proc. Symp. Networked Systems Design and Implementation **(NSDI** '07), Apr. 2007**.**

[4] C.A. Waldspurger, **"Memory Resource Management in VMware ESX Server,"** Proc. Symp. Operating Systems Design and Implementation (OSDI '02), Aug. 2002.

[5] G. Chen, H. Wenbo, J. Liu, S. Nath, L. Rigas, L.Xiao, and F. Zhao, **"Energy-Aware Server**

[6] **Provisioning and Load Dispatching for Connection-Intensive Internet Services,"** Proc. USENIX Symp. Networked Systems Design andImplementation (NSDI '08), Apr. 2008.

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

A Cloud Assisted Zigbee-Based Zoo-Animal Health Monitoring System Using Big Data and IoT services

C. SIVAKUMAR[1], S. RAJA[2], R. ISWARYA[3], J. NISHAMUGI[4]
[1]Assistant Professor, [2,3,4]Student
[1,2,3,4]Department of Compute Science & Engineering
Vinayaka Mission's Kirupananda Variyar Engineering College, Salem, Tamil Nadu, India

ABSTRACT

The system entitled has been developed to assisted through monitoring system with the help of big data and IOT services. Mainly zoo-animal and its health monitoring is one of the prime facts for the social-economic development of a nation. This necessitates monitoring the wellbeing of the animals and thereby increasing the value of the country. Advancement in wearable biosensors and wireless communication technologies help in ubiquitous health monitoring of animals. In this system provides the IOT services cloud assisted to support zigbee-based zoo-animal health and diet. The system gives entire zoo-animal detection plays a vital role in day-to-day life. It is important to detect the presence of animals entering into the human living, since it causes damage to life of people living. It is important to safe guard the life of human by detecting the presence of animal and take necessary actions to safeguard human life. It is also equally important to save the animals. In order to overcome the above drawback a warning system must be developed. This system focuses of PIR (Passive Infra Red Sensor) which senses the presence of zoo-animal. The microcontroller is the heart of the system. It controls every component of the system. The LCD monitor displays if the zoo-animal has been detected. Buzzer is used for alerting. With the help of GSM the warning message is sent to zoo-administrator. The Zoo-Animals main parameters like body temperature, heart rate and position tracking are acquired using respective sensors. The collected data are transmitted wirelessly over internet and are stored in a cloud server. A cloud database results are taken future reference using IOT technology. The system also alerts the admin/care takers at the critical conditions. The records that reflect the physical condition of animals collected in the cloud database helps veterinary doctors to provide effective treatment.

Index Terms – Health Monitoring System, IoT Services.

1. INTRODUCTION

Researches regarding animal detection is important field to numerous applications. Many algorithms and methods are developed by human being in order to better understanding on animal behaviour. Besides, these applications also can act as a warning system to human being from intrusion of dangerous wild animal for early precaution measures. These applications can be narrowed down to three main branches, namely detection, tracking and identification of animal

The animal tracking is the main topic in monitoring animal locomotive behaviour and its interaction with the environment. With the technology of sensor, radio frequency identification (RFID), and global positioning system (GPS), one of the applications is the development of new zoological systems for animal tracing ability, identification, and antitheft for the management and security of animal in zoo. By tracking the animal movements, it helps human to better understanding on living creatures on earth, especially on how the animal interacts with its environment.

The system focuses the various activities relating to animal health monitoring system such as:

- Temperature sensor module
- Humidity sensor module
- Heart rate sensor module
- Rumination sensor module
- Zigbee module
- Report processing module

2. EXISTING SYSTEM

System survey is a general term that refers an orderly, structural process for identifying and solving problems. System analysis is a process, life cycle methodology, since it relates to four significant phase in the life cycle of all business information systems: Study, Design, Development and Operation. The Definition of system analysis includes not only the process of analysis but also synthesis, which is the process of putting part together to form a new whole.

In the existing system is designed to manage the wildlife from fatal accidents in transport environments. The detector circuit on both sides of the route, detects the entry or presence of a wildlife animal which is about to cross the road. The detector circuit passes the information to the control unit by means of short-range communication. The controller section comprising of a microprocessor or microcontroller manages the traffic according to the signal received, by showing a red or stop signal on both directions. Once the animal has crossed the road the controller gives ready to go warning signal/sound is generated to scare away the creatures.

Existing system focuses the animal stays or is present in

the route even after the warning signal, which may be due to a prey attack or injury, a message is passed to a security personal. The security person after receiving the information takes the necessary according to the scenario. Thus, the animal and mankind both can be benefitted. If we are doing the system manually, so many minor errors will occur. Error detection in the previous entries made and data cross verification is another important function. These are done manually, and it would take time.

The existing system is carried out manually. There are lot of difficulties in the manual maintenance of the illegal records and offense reports. The existing system is carried out manually. There are lot of difficulties in the manual maintenance of the immoral records and fault reports.

DRAWBACKS OF EXISTING SYSTEM

- No API for Text and audio to save animal health.
- Existing system is time consuming and not very user friendly.
- The complaint brought by a person hailing from a family beyond poverty, the existing system shows ignorance for the same.
- No sensor or exact hardware controls are used in existing system.
- As we all know, a covered truth, fake reports play an important role in the existing system.
- In most of the hazard cases, the innocent is accused in the existing system.

3. PROPOSED SYSTEM

PIR sensors in target payload and movement type identification. The proposed system combines the cloud computing and embedded technology with the Zigbee based wireless communication technology, this system deals with a health monitoring and tracking system for animals using cloud service. This device tracks the animal's space and also measures the animal physiological signal by using Zigbee transceiver, IoT services and GPS.

The proposed system addresses the problem of target detection and classification using seismic and PIR sensors that monitor the infiltration of humans, light vehicles and domestic animals for border security. The major contributions of the system are as follows:

- Formulation of a hierarchical structure for target detection and classification.
- Experimental validation of the SDF-based feature extraction method on seismic and PIR sensor data.
- Performance evaluation of using seismic

In this proposed system to neck of animal this light weight designed system is attached such that temperature sensor will be very close to the body of that animal for IoT service. Thus, body temperature is sensed and sends to microcontroller properly.GPS modem will receive string from satellites and send to microcontroller. Then microcontroller will extract latitude and longitude information from string and send it to the GSM modem. After receiving the SMS forest officer will come to know the body temperature and location information. It is possible to locate exact geographical position of animal with the help of Google map. When latitude and longitude information is known, after entering this information on the Google map can locate it by using internet.

The objectives of this system is to demonstrate registration of pasture time in a specific area (a strip with new grass) using a ZigBee-based wireless sensor network and single hop connectivity. Another objective was to prove two extensions: an area extension where knowledge about animal presence in a limited area is used to predict animal presence in a larger extended area. The other extension aims at determining the whole herd presence based on registration of a subset of tagged animals. Yet another objective was to solve a specific problem regarding packet loss using data post processing. Each node in the network was programmed to transmit data when located within communication range of a gateway in the area with new grass. The principle is single hop connectivity that is the gateway only registers presence when a specific node is within the communication range and actively participates in handshaking communication.

In this research, cloud server, IoT service, multi hop connectivity as used in modern communication networks was not utilized. As the area defined by the communication range does not necessarily cover the same area as the new grass strip, an area-based correction factor was applied to the measured time in the gateway connectivity area to estimate the total pasture time in the new grass strip.

The application domain is a paddock in which wish to monitor the state of the animals and the landscape, and our test site is at Belmont near Rock Hampton. Several Fleck 1cs are inter-faced to digital weigh-bridges (via an RS232 serial link), as well as water trough flow meters (via an analog input).

Some fleck extension board that interfaces with up to 5 soil moisture sensors allowing measurement of the vertical moisture profile in the ground. The mobile component of this network is 20 Fleck 2s which are worn by the animals. All these flecks are connected to a central facility via a pair of Fleck 1cs acting as gateways to a PC which provides a route to the Internet via an ISDN link. The long hop from the paddock to the central farm building is achieved using high-gain antennas. The initial testing occurred before the network link was established. The Fleck 2s were programmed to write all sensor data onto the flash memory card in plain text format. However, the Flecks were also programmed to broadcast their identity at regular intervals and to record all received broad-cast messages. This allows us to build up and analyze connectivity information over time. The aim is to shortly deploy an application that relies heavily on the radio and does not require the flash memory card to be physically removed as often as it needs to in the current setup.

The approach is to deploy a wireless mobile sensor and actuation network, which is capable of estimating the dynamic states of bulls, and performing real time actuation on the bulls from location and velocity observations. As it is a challenging task to implement a real-world mobile sensor/actuation. The various functional units present in the proposed system are: the sensor unit, Arduino Uno board, GSM/GPRS/GPS module, cloud server and mobile App.

- **Sensor Unit:** Temperature and heart rate sensors are interface with the processing board that acquires the body temperature and the heart rate of the animal under test.
- **Arduino:** Uno board is used for processing the sensed data in which a processor is built in.
- **GSM/GPRS/GPS** module is also interfaced with the Arduino Uno board that helps in tracking the animal and also transmits the data wirelessly to the cloud. The GPRS enables to send SMS to the farmer/zoo officials at any critical situations.

Our proposed system consists cloud system is discussed, Main components of the hardware system

- Microcontroller based motherboard with regulated power supply.
- Power Supply
- Heart Rate
- Printed Circuit Board (PCB)
- Temperature Sensor
- Humidity Sensor
- Liquid Crystal Display (LCD)
- GPS receiver for location information.
- GSM modem/mobile phone for remote communication.

ADVANTAGES OF PROPOSED SYSTEM

- Software and Hardware controls are used for this system.
- Sensors signals as processing to watching the animal health.
- Reducing the hazard disorder for animals.
- No confidently and anonymity of animal health issues.
- Provides strengthening and enhancing the protected area network.
- Provides effective management of animal health by IoT and cloud storage.
- Provides monitoring and research the animal health.
- Provides human resource development and personnel planning.
- Provides animal health awareness and hazard situation.

The proposed system is designed to overcome all the disadvantages of the existing system.

4. MODULES

"A Cloud Assisted Zigbee-Based Zoo-Animal Health Monitoring System Using Big Data and IoT Services" has the following modules:

- Temperature Sensor Module
- Humidity Sensor Module
- Heart Rate Sensor Module
- Rumination Sensor Module
- Zigbee Module
- Report Processing Module

TEMPERATURE SENSOR MODULE

This module gives the Core Body Temperature (CBT) range of the animal in which metabolism functions without modification, termed the thermo-neutral zone. Typically, core body temperature is higher than ambient temperature to ensure that heat generated by metabolism flows out to the environment. Deviation outside of this range which is relatively narrow leads to increases in resting metabolism, modifications to the biochemistry and cellular physiology as well as the behavior of the animal. A healthy adult animal body temperature range is approximately range is fixed and if the animal body temperature is over that this range then can called the cow is not healthy.

HUMIDITY SENSOR MODULE

This module gives environmental parameters are affected the performance and health of the animal both directly and indirectly. The environmental factor consists of air temperature, air movement, humidity, and radiation heat.

It mainly considers the environment temperature and humidity. Based on these parameters. To calculated the thermal humidity index (THI) and also analyze the stress level of the animal.

HEART RATE SENSOR MODULE

The heart rate sensor module covers the most important parameter in the health assessment. The adult healthy animal has a heart rate between 48 and 84 beats per minute. The variation in heart rate normally reflects the stress, anticipation, movement, exertion, and various diseases. Basically, the heart rate measurement is an indirect method.

RUMINATION SENSOR MODULE

The rumination module is a direct indicator of animal wellbeing and health and also important part of the process by which animal digest food. According to veterinarians, the rumination is a function of what the animal has eaten and how well He/She has been able to rest. Normally, animals spend about one third of a day (9-10hours) in ruminating. The changes of rumination module are indicating the disease such as mastitis, metabolic calving disease, food digestion, etc.

ZIGBEE MODULE

Zigbee module is used to communicate animal telemedicine is one hot application in the area of wireless sensor network. The communication between the sensor module and sink module is performed from side to side a zigbee module. The zigbee module working on the 2.4

GHz band, but is data transmits and receives serially through UART (universal asynchronous receiver transmitter). It covers also serially data transfer between zigbee coordinator and graphical user interface PC

REPORT PROCESSING MODULE

This module is used to report processing of the data is tested for the real time monitoring of physiological parameters such as body temperature, rumination, and heart rate as well as monitor the surrounding humidity, and temperature.

5. IMPLEMENTATION

Implementation is the stage, which is crucial in the life cycle of the new system designed. **"A Cloud Assisted Zigbee-Based Zoo-Animal Health Monitoring System Using Big Data and IoT Services"** is used to implement the experimental validation of animal health monitoring system (AHMS) which is capable to the measuring of body temperature, rumination, and heart rate parameters with environmental parameters (surrounding temperature and humidity). The system is based on the IEEE 1451.2, IEEE 802.15.4, and IEEE1451.1 standards. The PIC18F4550 microcontroller and XBee-PRO S2 module were used to the development of AHM system. The four-sensor module such as body temperature, heart rate, surrounding humidity and temperature and rumination has been successfully developed.

They measuring parameters will be helpful to analyze the animal disease or health condition of the animal. The proposed system designed front panel. The front panel of the AHM system handles functions of the measuring parameters such as settings the time interval, start but-ton (ON), OFF, data saves for the access memory of the PC or in the data base, and a digital and a graphical output. Here the developed GUI module can perform for four sensing module and display the seven valuable physiological and behavioural parameters. The USB slot of the PC is present the 100mA at 5V and it does not require any external power source in the sink module during the experiments. The power consumptions in the AHM system are depend only on the wireless sensing modules. During experiment, the 11.1V battery (rated 350mAh) is used. Each sensor module could be run incessantly for 60 hours without necessitate recharge.

6. CONCLUSION

A prototype of an animal health monitoring system is presented. The prototype system consists of the sensor module and sinks module. This system may be implemented in the wild life sanctuaries in addition to this fire accidents in the forest also be stopped by alerting the concern persons. Can implement it in the houses where there are pet animals. It will be specifically targets health monitoring during races, animal location and tracking applications.

The system has been developed ergonomically with the reference of the animal, the veterinary staff and primary user of the device. The following points are followed by the designing of the system in terms of the reduction of environmental factors, such as, the module is protective covering of PVC (Polyvinyl chloride) to shield it from rain and insects as well as the design of the casing for the collar to be threaded through, minimum noise is achieved in the case of the developed multilayer circuit board which includes a ground plane, and sensor and its associated circuitry are connected through wires with grounding connection

This technology presents very high low power consumption, low Complexity and time domain resolution. In the heart rate sensing module used the Rs232 transmitter and the developed module has been transmitting data only up to 2 meters. They will need the modification of the heart rate sensor module and could be increased the transmission range.

The prototype designed for screening the vital parameter of the animal under test is simple and user-friendly. Many livestock holders and veterinary doctors will be benefited on using this prototype. They could take care of the animals and monitor them at any time and from anywhere. As a future work wireless sensors could be used to make the system still more comfort for the animals under monitoring.

7. REFERENCES

[1] Alvaro A. Cardenas (2017). Keynote: Security and Privacy in the Age of IoT. In Proceedings of CyberW'17, Dallas, TX, USA.

[2] Anuj Kumar and Gerhard P. Hancke, (2015). A Zigbee-Based Animal Health Monitoring System. IEEE Sensors Journal, Vol. 15, pp. 610 – 617.

[3] Anushka Patil, Chetana Pawar, Neha Patil, Rohini Tambe (2015). Smart health monitoring system for animals. Green Computing and Internet of Things (ICGCIoT), International Conference.

[4] Accelerometer ADXL available at:https://images-na.ssl-images-amazon.com/images/I/61F5gkx2PGL._SX342_.jpg

[5] ADXL 335, www.analog.com.

[6] Bangar Y, Khan TA, Dohare AK, Kolekar DV, Wakchaure Nand Singh B (2013). Analysis of morbidity and mortality rates in cattle in Pune division of Maharashtra state. Vet World, pp. 512- 515.

[7] B. Wietrzyk and M. Radenkovic, (2009). Enabling large scale ad hoc animal welfare monitoring. 5th Int. Conf. on Wireless and Mobile Communication (ICWMC 2009), Cannes/La Bocca, French Riviera, France, IEEE Computer Society.

[8] B. Wietrzyk, M. Radenkovic, and I. Kostadinov (2008). practical MANETs for pervasive cattle monitoring. Proc. of the 7th Int.Conf. On Networking, Cancun, Mexico.

[9] B. Wietrzyk and M. Radenkovic, "Energy Efficiency in the Mobile Ad Hoc Networking Approach to Monitoring Farm Animals" Proceedings. of The Sixth International Conference on Networking (ICN 2007), Martinique, French Caribbean, 2007

[10] Chao HsiHuang, PinYin Shen, Yueh Cheng Huang (2015). IoT based physiological and environmental monitoring

system in animal shelter. International Conference on Ubiquitous and Future Networks

[11] E. S. Nadimi, R. N. Jorgensen, V.B. Vidal, and S. Christensen. (2012) Monitoring and classifying animal behavior using zigbee based Electronic copy available at: https://ssrn.com/abstract=3315327 Special Issue based on proceedings of 4th International Conference on Cyber Security (ICCS) 2018 INTERNATIONAL JOURNAL OF ADVANCED STUDIES OF SCIENTIFIC RESEARCH (IJASSR) ISSN 2460 4010 ABSTRACTED & INDEXED IN ELSEVIER-SSRN 30 mobile ad hoc wireless sensor networks and artificial neural networks. Computers and Electronics in Agriculture, ACM, vol. 82, pp. 44- 54.

[12] E. S. Nadimi, H. T. Sogaard, T. Bak, and F.W. Oudshoorn (2008). Zigbeebased wireless sensor networks for monitoring animal presence and pasture time in a strip of new grass. Computers and Electronics in Agriculture, ACM, vol. 61, pp. 79-87.

[13] E. S. Nadimi and H. T. Sogaard (2009). Observer kalman filter identification and multiple model adaptive estimation technique for classifying animal behaviour using wireless sensor networks. Computers and Electronics in Agriculture, ACM, vol. 68, pp. 9-17. [14] E. Lindgren, "Validation of rumination measurement equipment and the role of rumination in dairy cow time budgets," Thesis, Swedish University of Agriculture Sci., 2009.

[14] Heartbeat sensor available at: https://probots.co.in/images/large/ArduinoPulseSensor_01_LRG.jpg

[15] H. Hopster and H. J. Blokhuis (1994). Validation of a heart-rate monitor for measuring a stress response in dairy cows. Canadian J. of Animal Sci., pp. 465-474.

[16] Hugo Filipe Lopes and Nuno Borges Carvalho (2016). Livestock low power monitoring system. IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet).

[17] Jacky S. L. Tings, K. Kwok, W. B. Lee, Albert H. C. Tsang and Benny C. F. Cheung (2007).

[18] A Dynamic RFID-Based Mobile Monitoring System in Animal Care Management Over a Wireless Network. International Conference on Wireless Communications, Networking and Mobile Computing.

[19] Ji-De Huang and Han-ChuanHsieh (2013). Design of Gateway for Monitoring System in IoT Networks. IEEE International Conference on and IEEE Cyber, Physical and Social Computing.

[20] J. I. Huircan, C. Munoz, H. Young, L. V. Dossow, J. Bustos, G. Vivallo, and M. Toneatti (2010). Zigbee based wireless sensor network localization for cattle monitoring in grazing fields. Computers and Electronics in Agriculture, vol. 74, pp. 258-264.

[21] K. R. Lovett, J. M. Pacheco, C. Packer, and L.L. Rodriguez (2009). Detection of foot and mouth disease virus infected cattle using infrared thermography. The Veterinary J., vol. 180, pp. 317- 324.

❑❑❑

Choosing the Right Sensors to Ensure Effective Transmitter Localization

R. UMAMAHESWARI[1], C. MANIKANDAN[2], C. POONGODI[3], P. SRI JANANI[4]
[1]Assistant Professor, [2,3,4]Student
[1,2,3,4]Department of Compute Science & Engineering
Annapoorna Engineering College, Salem, Tamil Nadu, India

ABSTRACT

Using a distributed network of sensors, we tackle the issue of locating an unlicensed emitter. Our goal is to create methods for efficiently performing transmitter localization, where efficiency is measured in terms of the quantity of sensors utilized *to locate. An essential issue that occurs in many crucial applications, such as checking shared spectrum systems for any* unauthorized *users, is the localization of illicit transmitters. It is crucial to develop methods that reduce the energy resources of the sensors because the localization of transmitters is typically based on observations from a deployed set of sensors with constrained resources. In this study, we develop a secure data transmittion using the rc4* algorithm. *By using this* algorithm, *we encrypt the location information of the user who is sharing the information in the* network. *Doing this we can assure that the location information is hidden even in the server where we maintain our data If a user is searching for a information he will get only the details and not the location information of the owner who has shared the information in the network. By doing this we can assure that the information about the location of the owner is kept safe and secure inthe server.*

Index Terms - Approximation Algorithms, Cognitive Radio, Energy Efficiency, Radio Spectrum Management.

1. INTRODUCTION

Wireless transmitter localization via analysis of the received signal from multiple receivers or sensors is an important problem. While the problem has been widely explored, it exposes new challenges in many emerging applications due to the constraints of the application. In this work, we are specifically interested in a distributed mon- itoring system where a set of distributed RF sensors are tasked to detect and localize transmitters. These transmitters could be of various type. For example, in certain spectrum allocation scenarios, unknown primary transmitters need to be detected/localized, or in spectrum patrolling scenarios, unauthorized transmitters need to be detected/localized [1].

Recent work has explored new approaches for such monitor- ing where the RF sensors are crowdsourced, perhaps using various low-cost spectrum sensing platforms [2], [3]. The crowdsourcing deploys a large number of sensors. Fine grained spectrum sensing is implemented by creating suitable incentive mechanisms [2], [4].

Crowdsourcing makes the sensing cost-conscious. The cost here could be incentivization cost, cost of power, backhaul bandwidth on the part of the spectrum owner or the opportu- nity cost – being low-cost platform, the sensors may be able to only sense smaller spectrum bands at a time. Thus, involving only a small number of sensors or sensors with low overall cost budget (for a suitable cost model) for sufficiently accurate localization performance is critical. Prior work that discusses sensor selection in this context only presents heuristics without any performance guarantees [2].

We do not use geometric approaches which rely on hard-to-model mapping of received power to distance. Instead, we use a hypothesis-driven, Bayesian approach for localiza- tion [5]. We focus on the optimization problem of selecting a certain number of sensors from among the deployed sensors such that an appropriately defined objective of localization accuracy is maximized. This optimization problem can also be used to solve the dual problem of selecting a minimum number of sensors (or sensors with the minimum total cost budget) to ensure at least a given localization accuracy. We adopt the framework of a hypothesis-driven localization approach wherein each hypothesis represents a configuration (location, power, etc.) of the potential transmitters and then the local- ization is equivalent to determining the most-likely prevailing hypothesis. See Figure 1. The hypothesis-driven framework does not require an assumption of a propagation model, and works for arbitrary signal propagation characteristics.

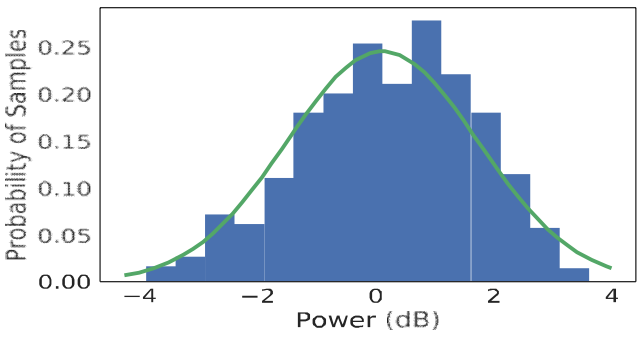


Fig. 2: Distribution of the received power from a

transmitter at an RTL-SDR sensor, and the Gaussian fit (green line) of the observed distribution. Thetransmitter and the sensor are kept in the corridor of a large building at thesame height, 10m apart from each other

The (overall) *probability of error* for a given set of sensors **T** is given by:

P (T) = mΣ−1 P (MAP(x) i|H)P (H) (5)

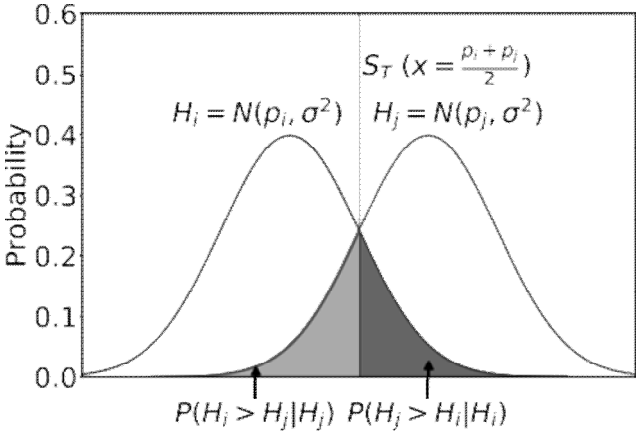


Fig. 3. Classification of a data point between two Gaussians using a threshold.

These assumptions using our own sensor in the wild (as shown in Figure 2). These assumptions have also been made by multiple prior studies [12], [13]. The covariance matrix remains same across hypotheses, since the correlation and noise are properties of the sensors. The means pi can be error different, as different power values are received by the sensors depending on the location of the transmitter.

Localization Accuracy Function, Oacc(**T**)*:* To facilitate a greedy approximation solution, we formulate our sensor selection as a maximization problem—and thus, define a corresponding maximization objective. In particular, we define the localization accuracy Oacc(T) as 1 − Perr(T).

Properties of MAP Algorithm

To explain our sensor selection algorithm, we first need to explain a few properties of MAP algorithm. Assume that there on the above equation Eqn. 5, we get the expression for are two hypotheses Hi and Hj , with distributions (pi, σ2) (**T**) as and (pj, σ2) as well as priors P (Hi) and P (Hj) respectively, where pi, pj ∈ R. Without loss of generality, we assume

O (T) = 1 − P
(T) = P (MAP(x) = i|H)P (H)

that pi < pj. In this case, given a data point X , the MAP

acc err (shown in Figure 3). If X ≤ ST , then MAP classifies X as Hi, i.e. MAP (X) = i, otherwise it classifies X as Hj, i.e. M AP (X) = j. Note that because this is a stochastic decision, there will always be some probability of classifica- tion error, depending on the value of ST. The MAP algorithm uses the threshold value of ST = pi+pj , and it is well-known with minimum probability of error Perr(T) (or maximum 2) that this value of ST provides the lowest probability of localization accuracy Oacc(T)), under the constraint that T| |

classification error. Formally, we write this as:

is at most a given budget *B*. Formally, the formulation is:

Hj pi + pj

P (Hi)

Maximize Oacc(T) subject to |T|≤ B. (7)

X ≷ 2 + log P (H) (8)

The above formulation implicitly assumes a uniform cost for each sensor; we generalize our techniques to handle non- uniform sensor costs (see §III-G).

We show that the above *LSS* problem is NP-hard, via reduction from the well-known maximum-coverage problem (Appendix A). Thus, we develop approximation algorithms below; in particular, our focus is on developing greedy approx-imation algorithms. The key challenge lies in showing that the objective function satisfies certain desired properties that ensure the approximability of the algorithm.

We now explain the case for multidimensional distributions, where H_i and H_j are given by $(\mathbf{p}_i, \Sigma)$ and $(\mathbf{p}_j, \Sigma)$ respectively

($\mathbf{p}_i, \mathbf{p}_j \in R, \Sigma \in R \times R$). In this case, the classification of a given data vector can be done by comparing with a

hyperplane. However, this problem of classification between distributions with multiple dimensions can be *reduced* to classification between distributions with single dimensions, using the following theorem:

Theorem 1: Given the hypotheses Hi ∼ N (**p**i, Σ) and

$H_j \sim N(\mathbf{p}_j, \Sigma)$, *a data vector* $\mathbf{x} = [x_1 \ldots x_n]$ *can be*

classified by applying the following threshold test:

B. Transmitter and Sensor Model $\mathbf{x}^T \Sigma^{-1}(\mathbf{p} \quad \mathbf{p})$
Hj $\frac{1}{}(\mathbf{p} + \mathbf{p})^T \Sigma^{-1}(\mathbf{p} \quad \mathbf{p}) + \log \frac{P(Hi)}{}$

We now formally define the assumptions that would allow us to ensure the approximability of our algorithm.

First, we assume that the joint probability distribution (JPD) follow

$$j - i \gtrless \quad i \quad j$$

$$Hi$$

$$j - i \quad P(H) \qquad (9)$$

a joint Gaussian distribution with the means (pi, Σ) for all hypotheses Hi, Ai = 0, . . . , m − 1. We empirically verify.

2. RELATED WORK

Indoor Localization

Indoor localization has been a topic of interest for a long time [24], [25]. Our technique for the hypothesis-based frame- work utilizes the fingerprinting technique [26] that has been discussed in earlier works. The work [27] fuses IMU sensors and WiFi RSSI measurements to improve the accuracy of indoor localization. Similar techniques have been used usingsound waves too [28].

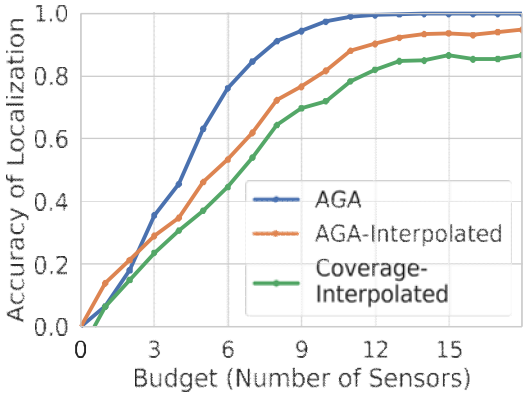


Fig. 13. Performance of AGA and Coverage algorithms when half the JPD's are obtained from empirical measurement, and the other half is obtained by interpolation

The key observation is that, for a given hypothesis Hi, the probability distribution of observations at a sensor s depends only on the configuration of transmitters in Hi that within a distance of R of s. I.e., for any observation xs at a sensor s, P (xs| Hi1) = P (xs H| i2) for any two hypotheses Hi1 and Hi2 that have the same configuration (locations and powers) for transmitters that are within a distance of R of s. The implication of the above observation(s) is that, for a given s, we can group the given hypotheses into equivalence classes based on the configuration of transmitters close to s, and to compute the benefit of a sensor s with AGA's iteration, we only need to compare pairs of equivalence classes (rather than the original hypotheses, which are exponentially many). The number of such equivalence classes is easily seen to be equal to GT where GR is the number of locations (grid cells) within R times the number of power levels, and T is the maximum number of transmitters possible/allowed within a range R of s (or any location). Thus, computation of benefit of s requires consideration of G2T pairs of equivalence classes. If we assume T to be a small constant, then the overall time complexity of AGA reduces to O(nB3) as before, and to O(nB) if we assume independence of sensor observations.

In our work, we have assumed the existence of only a single transmitter in the area under consideration. The rationale behind this assumption is that in many applications multi- ple concurrent transmitters do not exist due to the use of an effective multiple access protocol that avoids concurrent transmissions in the same neighborhood. Transmissions from far-away transmitters can be treated as noise.

Presence of Training Data

Our framework assumes that training data for each of the hypothesis is available. This training is usually expensive as it requires a lot of manual effort. While reducing training effort involved in utilizing MAP is not the primary focus of this work, we studied the performance of our techniques when we collected only half the original training data. We obtained the means of the rest of the joint probability distributions (JPD's) by linear interpolation. We then compared (Figure 13) the performance of AGA and Coverage algorithms with and without interpolations.

We observe that the performance of the algorithms do reduce on reducing the amount of training. The reduction in performance is highest (close to 18% at budget of 7) when the budgeted sensors is moderately high, but it reduces (around 8% at budget of 5) with further increases in the budget. While for clarity we do not show the reduction for the other techniques, this reduction in performance is observed for all the techniques, as they all depend on MAP for the final localization. We leave it to future work to investigate better interpolation techniques to enable more accurate localization.

Knowledge of Selected Sensors by Transmitters

In this work, we have assumed that the transmitters are unaware of the sensors that are selected. This is because our work is evaluated on the prior probabilities of each hypothesis being equal. If the transmitters are aware of the selected sen- sors, in certain types of applications (e.g., spectrum patrolling problems when the transmitters are unauthorized) they would try to evade the sensors by appearing at locations that are less closely monitored. This in turn would gradually change the prior probabilities, leading to a change in the subset of selected sensors. Studying the changing dynamics of how the unauthorized transmitters and selected sensors can react to changing priors if left for future work.

3. CONCLUSION

Choosing the right sensors for transmitter localization is critical for achieving high accuracy and robustness. In this paper, we presented a comprehensive survey and

evaluation of different sensor modalities and techniques for transmitter localization. We discussed the challenges and trade-offs involved in selecting the right sensors and proposed a sensor selection framework that considers factors such as cost, power consumption, range, and interference. We also proposed a framework for sensor fusion and machine learning to enhance the accuracy and reliability of localization. Our experimental results demonstrate the effectiveness of our proposed approach in achieving high localization accuracy and robustness while minimizing.

4. REFERENCES

[1] acharya, A. Chakraborty, S. R. Das, H. Gupta, and P. M. Djuric´, "Spectrum patrolling with crowdsourced spectrum sensors," IEEE Trans. Cogn. Commun. Netw., vol. 6, no. 1, pp. 271–281, Mar. 2020, doi: 10.1109/TCCN.2019.2939793.

[2] M. Khaledi et al., "Simultaneous power-based localization of transmit- ters for crowdsourced spectrum monitoring," in Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw. (MobiCom), Oct. 2017, pp. 235–247, doi: 10.1145/3117811.3117845.

[3] A. Nika et al., "Empirical validation of commodity spectrum monitor- ing," in Proc. 14th ACM Conf. Embedded Netw. Sensor Syst. CD-ROM (SenSys), Nov. 2016, pp. 96–108, doi: 10.1145/2994551.2994557.

❑❑❑

Efficient Authentication System for TransactionThrough Face Recognition Approach

DR. T. BUVANESWARI[1], T. ANITHA[2], P. KARTHICK[3], M. RAMANI[4]
[1]Professor, [2,3,4]Student
[1,2,3,4]Department of Compute Science & Engineering
Annapoorna Engineering College, Salem, Tamil Nadu, India

ABSTRACT

Most applications for computer security and privacy employ password-based authentication. Human error, such as selecting poor passwords and entering passwords incorrectly. Users typically choose passwords that are either brief or meaningful in order to make them easier to remember than random alphanumeric strings. People can access these applications anytime, anywhereand on a variety of devices thanks to the proliferation of online and mobile applications. A cutting-edge authentication system called Pass Matrix is currently in use to fend off shoulder surfing attacks. Pass Matrix provides no suggestion for attackers to figure out or narrow down the password even if they execute several camera-based attacks. It has a one-time valid login indicator and circulative horizontal and vertical bars encompassing the complete scope of pass-images. Nevertheless, this authentication mechanism is likewise not completely safe. One of the biometric authentication procedures has been implemented in our proposal. Here, A face is taken forthe authentication process and to improve security to ensure that a certain individual is physically present. This proves that the approach we have suggested achieves the highest level of authentication security.

Index Terms - E-Commerce, Debit Card, Credit Card, Lbph, Harrcascade, Face Recognition.

1. INTRODUCTION

In today's technologically advanced world, it's easy to hackers to get personal information of users so some People are afraid to use online transactions. This makes Securityas an important factor in the digital age Payments. Therefore, we propose a system to enhance security Transact online by providing a verycrucial verification process: OTP verification or by facial recognition.

During the transaction process the system will first verify the users face, if it matches then the transaction will be successfully completed and in case if the users face does not match then an otp will be sent to the user in order to successfully complete the transaction. The system, which we are able to propose, will try and lessen the variety of assaults on the time of creating virtual payments. Online transactions become vulnerable to loss or theft card, account theft, card forgery, fraud applications, multiple footprints and collusion merchants. In the case of account takeover, a card holder unknowingly offers his banking info to a fraudster and the fraudster then makes a replica card with the one's info. Among the colluding merchants, the employees of Merchants work with scammers. Suggestion system succeed in reducing all these frauds by capturing and verifying Real-time image of the cardholder. Biometric authentication is attracting a lot of attention due to unique to each individual. Several different biometric data authentications is fingerprint, hand geometry, iris, face and Palm. In this paper, we are using face recognition as it's the most popular, easily usable and widely acceptable. This system uses a computer system, a bank account to perform transactions and identify users. They provide PINs for security purposes. Use the right pin for access. But purchaser now no longer use right pin then now no longerbe verified. In manycases, debit or credit cards are lost when unauthorized users can access personal information such as passwords, phone numbers, shared birthday numbers. They easily guess the PIN. So we need to improve security such as strong passwords. But the authorized person easily use the password at that time by facial recognition technology to enhance the security and the user's information is authenticated.

2. LITERATURE REVIEW

Cancellable Biometric Filters for Face Recognition: In this article, we cover the issue of producing revocable biometric templates; features essential in implementing any biometric authentication system. We propose a new scheme encoding the training images used to synthesize a unique minimum mean correlation energy filter for biometric authentication. Theoretically, we show that transforming the training images with any random convolution before constructing the biometric filter does not change the obtained correlation output peak ratio, since that maintains authentication performance.

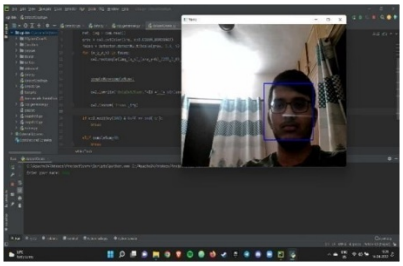
An Associate-predict Model for Face Recognition: Managing variations between parties is a major challenge in facial recognition. It's hard to know how appropriately measure the similarity between human faces in significantly different settings (e.g. pose, lighting, and expression). In this paper, we propose a new model, called "Associate-Predict" (AP) model, to address this issue. The predictive association model is built on top of a complementary common identity dataset, where each identity contains multiple images with large intraspecific variability.

Facial Feature Detection for Face Authentication: This paper presented an efficient algorithm for facial feature detection and a method for to automatically extract features for face authentication. The approach consists of combining different methods of region detection and feature extraction from face images. The Gabor filters, texture features, feature vector dimensionality reduction method are described.

Partial Face Recognition: An alignment free approach: Many methods have been developed for comprehensive face recognition with impressive performance. However, very few studies address the question of how to recognize an arbitrary image patch of anoverall face. In this paper we ad- dress this problem of partial face recognition. Part of a face often appears inunrestricted photography environments, especially when the face is captured by a surveillance camera or mobile device (e.g. mobile phone). The proposed approach uses a variable-size descriptor representing each face with a set of key point descriptors.

Enhancing User Authentication of Online Credit Card Payment: The popularity of online transactions over the past decade has resulted in the leakage of information about users. The security of the transaction process can be hacked more easily with advanced technologies. Biometric verification is seen as the key to improving security. In this article, a new process with face- match verification is proposed to improve the security of online payment system. Simulation of the online payment process is also created and then executed. The proposed new process is evaluated.

3. METHODOLOGY

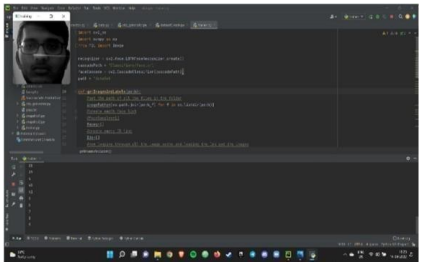


Module 1: Dataset Creator

In this module, we'll create a dataset that will hold photos of the user alongside the information they've provided. This module makesuse of a number of libraries, including cv2 for computer vision, numpy for image arrays, and sqlite3 for database dataset creation. Allof the photographs are stored in array format, which means that the pixels of the image are stored as an array. Account.db will be usedto store the information. The name of image stored will have id and the name details that are given by the user to the system in its first run.

The module will use the haar cascade algorithm to save the user's photographs using the cascade classifier function of the cv2 module. In this module the picture is taken as input using camera and stored in grayscale which is the format on which LBPH algorithm workson.

When the module is activated, the system will ask the user for their name and ID number. Further using cv2function the module will take input from camera, create a block where it detects face structure using harr cascade algorithm. When a face is detected, a block is created around it, which is subsequently saved in grayscale in the database. This module's code specifies the number of photos to be stored, which is currently set at 51.



Module 2: Trainer

This lesson involves training the dataset established in the previous module and exporting the trained data as a.yml file. This module also uses a lot of libraries which includes cv2 for image processing, os for accessing paths, numpy to modify and classify pixels of images that are stored in the form of an array. Other library used is pil whichis Python Imaging Library.

4. EXISTING SYSTEM

A basic multi-factor authentication setup consists of asking a user for their login and password (which they already know) and then verifying their identity with a second factor, such as an SMS message sent to their phone (something they have). That covers two authentication factors, but adding photo recognition to the mix offers an extra layer of protection without making the login procedure too complex or annoying for authorised users. Many banks employ picture recognition as part of their multi-factor authentication procedure so that their clients may securely access their accounts and authorise different financial activities. On the web, picture recognition authentication is great for combating phishing attempts in which a website imitates your bank's look and feel.

5. PRPOSED SYSTEM

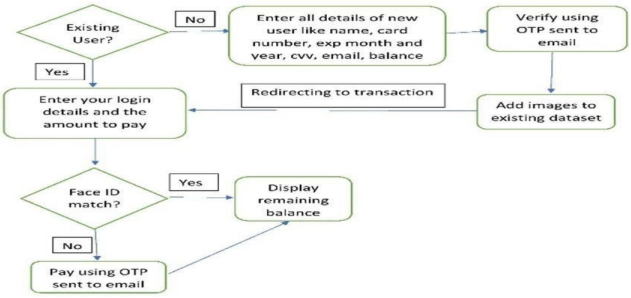


Figure 3. Data Flow Diagram

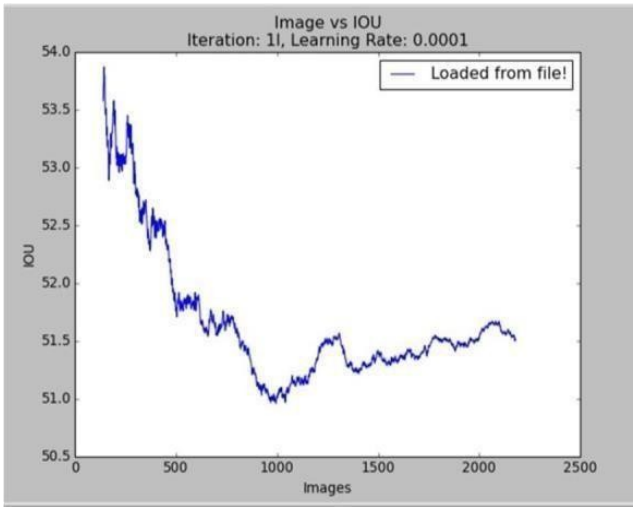
A data flow diagram (DFD) depicts the flow of data across an information system graphically. A DFD provides a high-level overview of the system without getting into too much detail.

The system's flow is as follows:

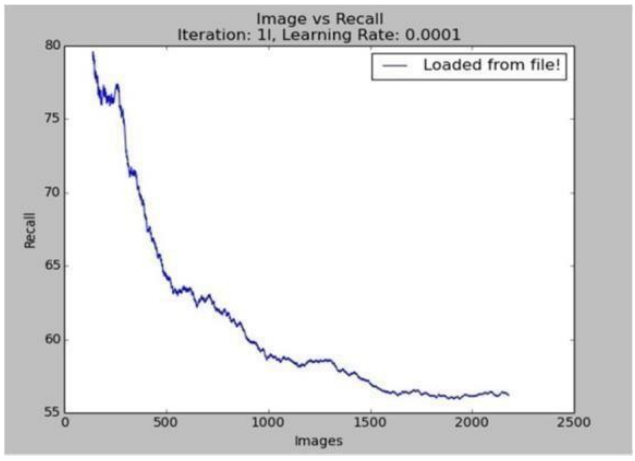
- The user inputs the desired amount and confirms it.
- Face verification is done by comparing input image with datasets.
- If face is successfully verified, payment is successful.
- If face verification fails, pin code is asked.
- If pin code is verified, then the payment is completed.
- If not verified, payment is declined.

Learning Rate (0.0001)

Learning rate is the training parameter that controls the size of weight and bias changes during learning. In the below graph x-axis stands for images whereas y-axis stands for IOU.

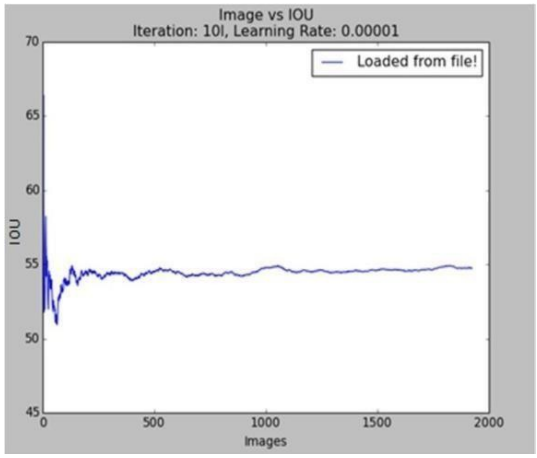


In the below graph x-axis stands for images and y-axis stands for Recall.



After conducting the above experiments with respect to training parameter Learning rate, the optimal value was found to be- Learning rate of 0.0001 detected objects, but with lesser accuracy in terms of classification.

Iterations: 800000



In the below graph x-axis stands for images whereas y-axis stands for IOU.

After conducting the above experiments with respect to training parameter Iterations, the optimal value was found to be- Number of training iterations:800000 For training below 800000 iterations, the network was found to predict fewer bounding boxes with less accuracy. Hence, the optimal number of training iterations was concluded to be as 800000.

6. CONCLUSION

In this paper, we proposed an authentication system based on face recognition technology for transaction authentication. The proposed system offers high levels of security and convenience for users. The experimental results showed that the proposed system achieved high accuracy in face recognition, making it suitable for real-world applications. Future work includes improving the system's performance on datasets with low-quality images and integrating the system with other biometric authentication methods.

7. ACKNOWLEDGMENTS

We would like to express my special thanks of gratitude to my Project guide "Dr. Ankita Karale" for her guidance and support that he gave to us for completing this project. The insights provided to us by her were very useful in successful completion of this project. I would also like to thank all the panel members for their thoughts and help they gave during reviews for improvement in the project.

8. REFERENCES

[1] Savvides, Marios, BVK Vijaya Kumar, and Pradeep K. Khosla. "Cancelable biometric filters for face recognition." Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. Vol. 3. IEEE, 2004.

[2] Q. Yin, X. Tang, and J. Sun, "An associate predict model for face recognition," in IEEE Conference on Computer Vision and Pattern Recognition, June 2011, pp. 497–504.

[3] R.S.Choras. "Facial feature detection for face

authentication," in the Proceeding of IEEE Conference on Cybernetics and Intelligent Systems., 2013, pp.112- 116.

[4] S.Liao, A. K. Jain, and S. Z. Li, "Partial face recognition: Alignment free approach," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no.

[5] pp. 1193–1205,2013.5. ICIIBMS 2015, Track1: Signal Processing, Computer Networks and Telecommunications, Okinawa, Japan 978-1-4799-8562- 3/15/$31.00 ©2015 IEEE Enhancing User Authentication of Online Credit Card Payment using Face Image Comparison with MPEG7-EdgeHistogram Descriptor.

[6] 2016 Online International Conference on Green Engineering and Technologies (IC-GET). A Robust and secure authentication mechanism in online banking.

[7] Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017.Credit Card Fraud Detection Based on Transaction Behavior.

[8] 2017 IEEE: Fast and Efficient Implementation of Convolutional Neural Networks on FPGA Abhinav Podili, Chi Zhang, Viktor Prasanna.

[9] 2017 20th International Conference of Computer and Information Technology (ICCIT), 22-24 December, 2017 Convolutional Neural Network Approach for Vision Based Student Recognition System.

[10] 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications Facial Expression Recognition via Deep Learning.

[11] Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication. Real time Implementation of Face Recognition System, Authors: Neel Ramakant Borkar; Sonia Kuwelkar.

[12] 2017 IEEE 2nd International Conference on Signal and Image Processing A Real-time Face Recognition System Based on the Improved LBPH Algorithm. Xue Mei Zhao, ChengBingWei.

❑❑❑

Increase The Security and Minimize The Privacy Risk in Cloud Storage

MEENA VIJAYAKUMAR[1], KOWSALYA SASIKUMAR[2], KAVITHA GOPAL[3], MALINI BALASUBRAMANI[4], ANJALI MUTHU[5]
[1]Professor, [2,3,4,5]Student
[1,2,3,4,5]Department of Compute Science & Engineering
AVS Engineering College, Salem, Tamil Nadu, India

ABSTRACT

Without having to worry about maintaining and storing their data locally, customers may benefit from high-quality programmes and services that are available instantly thanks to cloud storage Protecting the integrity of outsourced data in cloud computing is a difficult problem, especially for users with low computing resources, because users no longer physically possess the outsourced data. Additionally, users should not need to worry about checking the integrity of the cloud storage; they should just be able to utilise it as if it were local. A secret sharing group key management protocol (SSGK) is currently utilised to safeguard shared data and the communication process from unauthorised access In SSGK, a group key is used to encrypt shared data, and the group key is distributed using a secret sharing technique Its ability to provide public auditing for cloud storage is crucial so that consumers can employ a third- party auditor (TPA) to verify the accuracy of outsourced data and relax. In order to introduce a TPA properly and successfully, the auditing process shouldn't add any additional online hassles for consumers or vulnerabilities compromising user data privacy. In this research, we provide a private public auditing mechanism for a secure cloud storage system.

Index Terms – Cloud Storage, Privacy Risk.

1. INTRODUCTION

Cloud storage is an important branch of cloud computing, the purpose of which is to provide powerful and on-demand data services to operating users. Due to the low-cost and high-Outsource their data storage to professional cloud services providers (CSP), which buoys the rapid development of cloud storage and its relative techniques in recent years. However, as a new cutting-edge technology, cloud storage still faces many security challenges. One of the biggest concerns is how to determine whether a cloud storage system and its provider meet the legal expectations of customers for data security. This is mainly caused by the following reasons. First, performance of cloud storage, a growing number of organizations and individuals are tending to cloud users (data owners), who outsource their data in clouds, can no longer verify the integrity of their data via traditional techniques that are often employed in local storage scenarios. Second, CSPs, which suffer Byzantine failures occasionally, may choose to conceal the data errors from the data owners for their own self- interest. What is more severe, CSPs might neglect to keep or even deliberately delete rarely accessed data that belong to ordinary customers to save storage space.

Therefore, it is critical and significant to develop efficient auditing techniques to strengthen data owners′ trust and confidence in cloud storage, of which the core is how to effectively check data integrity remotely. So far, many solutions have been presented to overcome this problem, which can be generally divided into two categories: private auditing and public auditing. Private auditing is the initial model for remote checking of data integrity, in which the verification operation is performed directly between data owners and CSPs with relatively low cost. However, it cannot provide convincing audit results, as both owners and CSPs are often wary of each other. Moreover, it is not advisable for the users to carry out the audit frequently, since it would substantially increase the overhead that the users may not afford. Thus, Ateniese et al. first presented the public auditing scheme, in which the checking work is customarily done by an authorized third party auditor (TPA). Compared with the former, the latter can offer dependable auditing results and significantly reduce users′ unnecessary burden by introducing an independent TPA. In the public auditing, however, some vital problems as follows remain to be addressed or further gone.

2. INPUT AND OUTPUT DESIGN

The process of transforming a user- centered description of the input into a computer- based system is known as input design. This design is crucial to preventing mistakes in the data entry process and providing management with clear instructions for receiving the right information from the computerised system. It is accomplished by designing displays that are easy for users to utilise when entering big amounts of data. The purpose of input design is to make data entering simpler and error-free. The data entering panel is created such that any data manipulations are possible. Additionally, it offers record viewing capabilities.

3. IMPLEMENTATION

CLOUD STORAGE - Due to its financial benefits, data

outsourcing to cloud storage serversis becoming more and more popular among businesses and users. This effectively means that the data owner (client) transfers their data to a third-party cloud storage server, which is expected to faithfully store the data with it and provide it tothe owner when needed - probably for a price.

SIMPLY ARCHIVES - This issue aims to gather and verify evidence that the data stored by a user at remote cloud storage (also known as cloud storage archives or simply archives) has not been altered by the archive, guaranteeing the data's integrity. If cheating in this context means that the storage archive might destroy some of the data or might modify some of the data, then cloud archiving is not defrauding the owner.

PRIVACY PRESERVING - The protection of data privacy (DPP) has long been a key concern for cloud storage. The main issue in the public auditing is how to maintain user privacy while implementing a TPA. Although one method to reduce the privacy risk in cloud storage is to use data encryption before outsourcing, this method is unable to stop data leaking during the verification process. As a result, it's critical that the cloud auditing incorporate a privacy-preserving technique apart from data encryption.

SENTINELS - Contrary to the key-hash approach method, this scheme only allows the use of a single key, regardless of the size of the file or the quantity of files whose irretrievability it wants to verify. In contrast to the key-has technique, which required the archive to process the whole fileF for each protocol verification, the archive only needs to access a small piece of file F. If the proof has changed or removed a significant amount of F, it is highly likely that it has also suppressed a number of sentinels.

4. SYSTEM ANALYSIS

EXISTING SYSTEM - The purpose of cloud computing, which includes cloud storage, is to givecustomers who are utilising highly virtualized infrastructures access to robust, on-demand data outsourcing services. Due to cloud storage's low cost and great performance, an increasing number of businesses and people are choosing to outsource their data storage to specialised cloud services providers (CSP), which has fueled the technology'srecent rapid development. However, cloud storage still has a lot of security issues to overcome as a brand-new, cutting-edge technology. How to assess whether a cloud storage system and its provider meet the legal requirements of clients for data security is one of the main issues. The following factors are the main causes of this. First, users (data owners) of the cloud who outsource their data to clouds are no longer able to check the accuracy of their data using conventional methods, which are frequently used in local storage settings. Second, CSPs that occasionally experience Byzantine failures could decide against informing the data owners about the faults out of self-interest. Even worse, CSPs may purposefully erase rarely accessed data belonging to regular customers in order to preserve storage space. Therefore, it is critical and significant to develop efficient auditing techniques to strengthen data owners′ trust and confidence in cloud storage, of which the core is how to effectively check data integrity remotely.

PROPOSED SYSTEM - The foundation of a ground-breaking public auditing method for securecloud storage is the proposed system's use of dynamic hash tables (DHT), a new two- dimensional data structure located at a third parity auditor (TPA) to store the data property information for dynamic auditing. Differing from the existing works, the proposed scheme migrates the authorized information from the CSP to the TPA, and thereby significantly reduces the computational cost and communication overhead. Meanwhile, exploiting the structural advantages ofthe DHT, our scheme can also achieve higher updating efficiency than the state-of-the-art schemes. The homomorphic authenticator based on the public key and the random masking generated by the TPA are also combined in our scheme extension to support privacy preservation, and batch auditing is accomplished using the aggregate BLS signature technique.

5. SYSTEM REQUIREMENTS

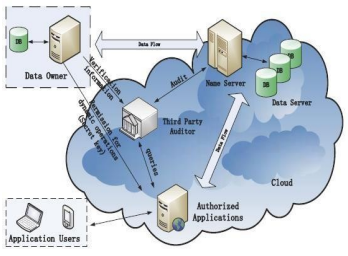
HARDWARE REQUIREMENTS

- System : Pentium Dual Core.
- Hard Disk : 250 GB.
- Monitor : 15’’ LED
- Input Devices : Keyboard, Mouse
- RAM : 2 GB.

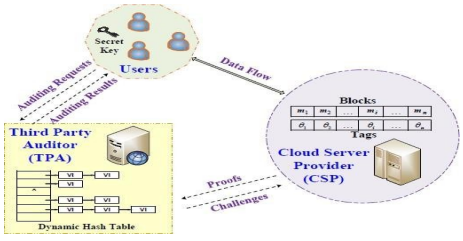
SOFTWARE REQUIREMENTS

- OperatingSystem : Windows XP,7.
- Language : PHP, JavaScript.
- Frontend : PHP 7.1.
- Backend : MYSQL 10.3

6. DESIGN SYSTEM DESIGN



SYSTEM ARCHITECTURE DIAGRAM



7. CONCLUSION

These days, cloud storage, which may provide on-demand outsourced data services for both businesses and people, is gaining popularity. However, one of the most significant barriers to its development is that users might not have complete confidence in the CSPs because it can be challenging to discern whether the CSPs meet their legal obligations for data security. To increase data owners' trust and confidence in cloud storage, it is essential and vital to establish effective auditing techniques Using dynamic hash tables (DHT), a brand-new two-dimensional data structure utilised to store the data property information for dynamic auditing, we are driven to provide a revolutionary public auditing system for safe cloud storage in this research. Differing from the existing works, our scheme migrates the auditing metadata excerpt the block tags from the CSP to the TPA, and thereby significantly reduces the computational cost and communication overhead. Meanwhile, exploiting the structural advantages of the DHT, our scheme can also achieve better performance than the state- of-the-art schemes in the updating phase. In addition, for privacy preservation, our scheme introduces a random masking provided by the TPA into the process of generating proof to blind the data information. In addition, our scheme further utilises the aggregate BLS signature technique frombilinear maps to carry out multiple auditing tasks concurrently. The principle behind this technique is to combine all of the signatures from various users on various data blocks into a single, condensed signature and verify it only once to minimise communication costs. As a result, creating a more effective scheme that includes diverse audit methodologies for different sorts of cloud data may be a new trend, and this is also the direction in which our future work will go.

8. REFERENCES

[1] H. Dewan and R. C. Hansdah. ″A Survey of Cloud Storage Facilities ″, Proc. 7th IEEE WorldCongress on Services, pp. 224-231, July 2011.

[2] C. Wang, Q. Wang, K. Ren, N. Cao and W. Lou. ″Toward Secure and Dependable Storage Services in Cloud Computing″, IEEETrans. Service Computing, vol. 5, no. 2, pp. 220-232, 2012.

[3] K. Ren, C. Wang and Q. Wang. "Security Challenges for the Public Cloud," IEEE InternetComputing, vol. 16, no. 1, pp. 69– 73, 2012.

[4] J. Ryoo, S. Rizvi, W. Aiken and J. Kissell. ″Cloud Security Auditing: Challenges and Emerging Approaches″, IEEE Security &Privacy,vol. 12, no. 6, pp. 68-74, 2014.

[5] C. Wang, K. Ren, W. Lou and J. Li. ″Toward Publicly Auditable Secure Cloud Data Storage Services″, IEEE network, vol. 24, no. 4, pp. 19-24,2010.

[6] Q. Wang, C. Wang, K. Ren, W. Lou and J. Li. ''Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing,'' IEEE Trans. on Parallel and Distributed Systems, vol. 22,no. 5, pp. 847-859, 2011.

[7] F. Sebé, J. Domingo-Ferrer, A. Martínez-Ballesté, Y. Deswarte and J.-J. Quisquater, "Efficient Remote Data Possession Checking in Critical Information Infrastructures,"IEEE Trans.Knowledge Data Eng., vol. 20, no. 8,pp. 1034- 1038, 2008.

[8] A. Juels and B.S. Kaliski Jr., "PoRs: Proofs ofRetrievability for Large Files," Proc. ACM Conf. Computer and Communications Security (CCS '07),pp. 584- 597, 2007.

[9] G. Ateniese, R.B. Peterson and D. Song, ''Provable Data Possession at Untrusted Stores,'' Proc. 14th ACM Conf. on Comput. andCommun Johns, R. Curtmola, J. Herring, L. Kissner, Z.. Security (CCS), pp. 598-609, 2007.

[10] K. Yang and X. Jia. ″Data Storage Auditing Service in Cloud Computing: Challenges, Methodsand Opportunities″. WorldWide Web, vol. 15, no. 4, pp. 409-428, 2012.

[11] C. Wang, Q. Wang, K. Ren and W. Lou, ''Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing,'' Proc. IEEEINFOCOM, pp. 1-9, 2010.

[12] C. Wang, S. M. Chow, Q. Wang, K. Ren and W. Lou, ″Privacy- Preserving Public Auditing for Secure Cloud Storage, ″ IEEE Trans. on Computers, vol. 62, no. 2, pp. 362-375, 2013.

[13] Y. Zhu, H. Hu, G. Ahn, and M. Yu, "Cooperative Provable Data Possession for Integrity Verification in Multi-Cloud Storage," IEEE Trans. Parallel and Distributed Systems, vol.23, no. 12, pp. 2231-2244, 2012.

[14] K. Yang and X. Jia, "An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing," IEEE Trans. on Parallel and Distributed Systems, vol. 24, no. 9, pp.1717-1726,2013.

[15] C. C. Erway, A. Küpçü, C. Papamanthou and R. Tamassia. "Dynamic Provable Data Possession," Proc. 16th ACM Conf. Computer and Comm. Security, pp. 213-222, 2009.

[16] Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu and S. S.Yau, "Dynamic Audit Services for Outsourced Storage in Clouds," IEEE Trans. onServices Computing, vol. 6, no. 2, pp. 227–238, 2013.

[17] D. Boneh, B. Lynn and H. Shacham, ''ShortSignatures from the Weil Pairing,'' Proc. ASIACRYPT, vol. 2248, LNCS, pp. 514-532, 2001.

[18] B. Wang, B. Li and H. Li. ″Panda: Public Auditing for Shared Data with Efficient User Revocation in the Cloud″, IEEE Trans. on ServiceComputing, vol. 8, no. 1, pp. 92-106, 2015.

[19] C. Liu, R. Ranjian, X. Zhang, C. Yang, D. Georgakopoulos and J. Chen. ″Public Auditing for Big Data Storage in Cloud Computing-- A Survey″,Proc. 16th IEEE International Conf. Computational Science and Engineering (CSE), pp. 1128-1135, 2013.

[20] C. Liu, J. Chen, L. T. Yang, X. Zhang, C. Yang, R. Ranjan and K. Ramamohanarao, ″ Authorized Public Auditing of Dynamic Big Data Storage on Cloud with Efficient Verifiable Fine-grained Updates, ″ IEEE Trans. on Parallel andDistributed Systems, vol.

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Source: Conscious Self Attention for Detecting IP Hijack

T. POORNACHANDAR[1], D. SOMASUNDARAM[2], M. UMAMAHESWARI[3], R. VANITHA[4]
[1]Assistant Professor, [2,3,4]Student
[1,2,3,4]Department of Compute Science & Engineering
Annapoorna Engineering College, Salem, Tamil Nadu, India

ABSTRACT

Man-in-the-middle attacks are caused when IP hijacking attempts divert traffic between endpoints through the attacker network. The primary basis for current detection techniques is AS-level path analysis, although attacks involving data-plane manipulations may show just geographical anomalies, maintain the AS-level route, or obfuscate the issue AS. Therefore, it is necessary to provide frameworks for data-plane analysis that look at the real paths that packets take. Here, we are going to transmit the packet to the destination using dynamic rooting technique. In this technique we are going to transmit the packet to the destination only after analyzing thenodes details. that is before we are transmitting the data server will verify the network for any ip hijack if it finds any hijack in the root offile transmition it will generate a new root for transmitting the packet, we not only do that we also send a dummy packet to the node which has got compromised, by doing this we will confuse attacker, we also transmit the packet securely by encrypting the packet. we implementthis in java swing.

Index Terms – IP Hijack, Self Attention.

1. INTRODUCTION

N RECENT years, there have been many reports of IP hijack attacks of nations and large companies, as more than 40% of the network operators reported that their organization had been a victim of a hijack in the past [1], [2]. In an IPhijack attack, the attacker diverts the traffic to its own network and then forwards it to the original destination, forming aman- in-the-middle (MITM) attack. This allows espionage, traffic manipulation, network penetration, and more. Since such attacks are hard to perform, they are mostly used by governments and large criminal organizations. Current solutions for IP hijack detection [3]–[5] are based on monitoring BGP routing announcements, and mostly detect changes due to change of origin AS, the first upstream providers, and some obvious malicious route changes. How- ever, hijack attacks are not limited to BGP manipulations and can also be performed with stealthier methods, e.g., by manip- ulating routing at the data plane in IXPs or inserting static entries to key ISPs. We show in this paper (see section VII-G), for the first time in the literature, an example of a suspected IP hijack attack that has no BGP signature and seem to be a result of BGP entry manipulation at the source ISP.

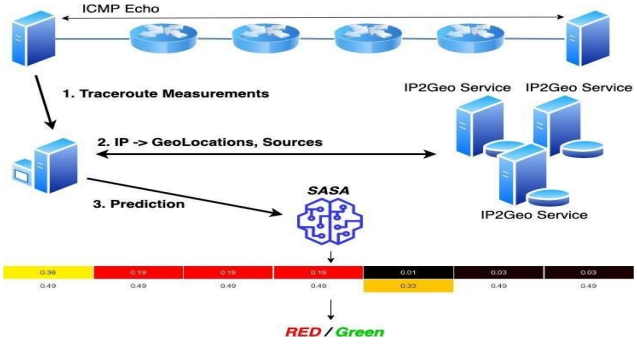


Fig. 1. Inference Information Flow

Thus, we need to develop IP hijack detection tools that examine the actual route packet traverse, which, during a data-plane attack, may not be the one announced in BGP.

It is important to note that routes may be deflected in unreasonable ways, also due to human error, not necessarily dueto malicious acts. Even such benign deflections expose traffic to MITM attacks, e.g., by traversing networks, which are involved in espionage and may lurk for interesting data. It is almost impossible to know the cause of a deflection without knowing the intent behind people's actions. Thus, we follow previous work [6] and throughout this paper we will use IP hijack and deflection interchangeably.

In this paper, we introduce a novel *data-plane* approach for IP hijack detection based on geographical data, using deep learning methods. We rely on the actual route that the packets traverse, obtain through traceroute measurements, rather than the path advertised by BGP [7]. Given the routers' IP addresses along the route, we obtain their geographic location and analyzethe route geography.

Our motivation for using geography for hijack detection is that although the primary routing decision criteria in BGP mostly derive from economic agreements between ASes, these contracts also reflect geographical and geopolitical constraints, as it is reflected in Figure 2. Hence, it is less likely to select paths that significantly deviate from the ideal direct geographical route, and certainly, it is unlikely for a routeto traverse unfriendly countries. There have been only a few works that examined using geographical data for characteriz- ing international detours in the Internet [7], [8] and some for visualization purposes [9]–[12]; however, none of them aimed at IP hijack detection.

Our data is obtained from large traceroute measurement campaigns from multiple agents around the world. The IP addresses are converted to geographical information using several geolocation services. Each route is labeled as hijacked or benign by three analysis algorithms: BGP Valley-free (VF) analysis that is based on Shavitt *et al.* [13] with manual cor- rections, geographical analysis, and ASN ownership analysis (see Sec. III).

The introduction of the Attention Mechanism [14] in deep learning has improved the success of various NLP models, by mapping the essential and relevant words from the input sentence and assign higher weights to these words, enhancing the accuracy of the output prediction. We build on the excellent results achieved for time series tasks and design a new layer that is based on the attention layer. Because there are many different services, with various levels of confidence [15], we design our layer to incorporate the data source confidence levelof each data sample. Thus, we called our layer Source- Aware Self-Attention (*SASA*). Just like any other parameter, our layeralso learns the confidence of each data source.

2. RELATED WORK

There are many different approaches for the detection of IP hijacking. We divide these approaches into three main cate- gories, based on the type of information they use: 1) Control- plane approaches [3]–[5] - also called passive solutions, these methods analyzed BGP routing information from a distributed set of BGP monitors and route collectors to detect anomalous behavior, 2) Data-plane approaches [2], [16], [17] - only relies on real-time data plane information that is obtained from multiple sensors that deploy active probing (pings/traceroutes). Some of these methods are based on analyzing IP TTL (Time to Live) or an increased RTT (round-trip delay time), and 3) Hybrid approaches [18]–[20] - these approaches use both control-plane and data-plane information and sometimes also use external databases to perform joint analysis.

There are several works that involved the use of geo-graphical data, mainly for visualization purposes [9]–[12]. Theodoridis *et al.* [11] introduced an unsupervised method that is based on three features related to the frequency of appearance and the geographic deviation of each intermediate AS towards a given destination country: the probability of an intermediate country appearance along a route toward a specific Origin- Country (CAP), the geographic length which is the ratio of the length of the path against the ideal direct path (CGL), and the Z-score of geographic lengths for all the intermediate countriesof a certain Origin-Country (CGLZ). Following their work, Papadopoulos *et al.* [12] pre- sented BGPfuse, a scheme for visualizing and exploring BGP path change anomalies, which used the three features which were mentioned before (CAP, CGL, CGLZ) with the addition of CAPZ, which is the Z-score of CAP. They used thesefeatures to quantify the degreeof the anomaly of each BGP hijacking event.

In 2004, Zhu and Wu [27] presented a systematic evaluationof the effect of class noise and attribute noise in machine learning, and analyze their impacts on the system performance. Following their work, many advances have been made in dealing with label noise [28]–[31]. However, as far as we knowwe are the first to present a deep learning method for dealing with unreliable data in multi-source datasets.

3. METHOD

In this section, we describe in detail the implementation of the *SASA* layer and the architecture of the deep neural networkwe designed for the classification.

SASA

Our Source-Aware Self-Attention layer is based on the Scaled Dot-Product Attention layer (*SDPA*), that was intro- duced by Vaswani *et al.* [14], which is a variant of the dot- product attention [40]. Let n be the number of elements in the As depicted in Table III, our LSTM architecture comprises four layers, not counting the input. Our input layer consists of 40 entities, which is the maximum length of routes in our datasets. In the case of a shorter route, we pad the remaining entities with 0s. As mentioned in Sec. III, in this work, we use two types of inputs: 1) coordinates, the latitudes and longitudes pairs divided by 90 and 180, correspondingly, such that each entity has a size of 2, and 2) countries, each country is indexed with a number between 0 and 246 (The datasets have 78 and 247 countries, respectively).

A sequence of coordinates or countries is fed into the first layer of thenetwork, which is an embedding layer (We omit the details). The next layer is the attention layer. In our experiments, we compare between different attention layers asdescribed in Sec. V-A. The inputs of each variation of the *SASA*layer also includes the source vectors, as described in Sec. V-

A. We use $d_k = d_v = d_s = d = 32$ such that the output remains with the same size as the input. The next layer is the Bidirectional-LSTM layer, which consists of 100 LSTMscells with a default configuration [37], and produces a 200- size output vector. Finally, our output layer is a single neuron witha Sigmoid activation function, which produces a value between0 and 1.

Training Specifications

The training of the networks is done by optimizing the *binary cross entropy* [42] cost function, which is a measure of the difference between the *Sigmoid* layer output and the true label of the sample. For the optimization process we use the *Adam* [43] gradient-based optimizer. Because our datasets are imbalanced, i.e., the 'RED' classes are mostly less than 1% of the datasets, we use a balanced generator that randomly samples the training set, such that in each batch, we have the same amount of 'RED' and 'GREEN' paths.

where TP , TN , FP and FN are the true positive, true

negative, false positive, and false negative, respectively. In our case the positive class corresponds with 'RED' routes,

False Alarm (FA), which is the False Positive Rate (*FPR*), defined as

We build and run our networks using the *Keras* [44] library with *Tensorflow* [45] as its back-end. We use 80% of the

$$FPR = \frac{FP}{FP + TN} \quad (6)$$

samples as a training set and 20% of the samples as a test set. We run our network for 60 epochs of the training set. We save the result, which achieves the best accuracy during the.

Detection Rate or **Recall** (Rc), which is defined by

TP

training process. $Rc = \quad . \quad (7)$

TP FN

4. EXPERIMENTS AND RESULTS

In this section, we report our experimental results. Since we have not found any previous work to compare our routes with, we compare our *SASA* layer with the regular Scaled Dot- Product Attention (*SDPA*) layer, as well as other variants we suggest. Since both *SDPA* and *SASA* present good attention performance, we will devote most of the evaluation section to the gain in accuracy of hijack detection obtained by the *SASA* layer.

Evaluation Criteria

Any classification system for anomalydetection (in our case in many cases, one would like to control the trade-off between the false alarm rate and the detection rate. Namely, set the accepted false alarm (which can be done here by settingthe threshold for the prediction score) and aim at the highest possible detection rate. Thus, we introduce two additional evaluation criteria:

AUC, which is the area under the ROC curve (a plot of the true positive rate (TPR) against the false pos-itive rate (FPR) at various thresholds), as displayed in Figures 4 and 5, and

True Positive Rate (TPR) as defined by

TP three

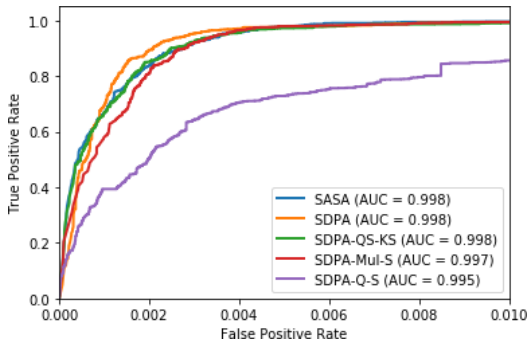


Fig. 4. ROC curves of models on 'Dataset A' - test set for FPR 0.01.

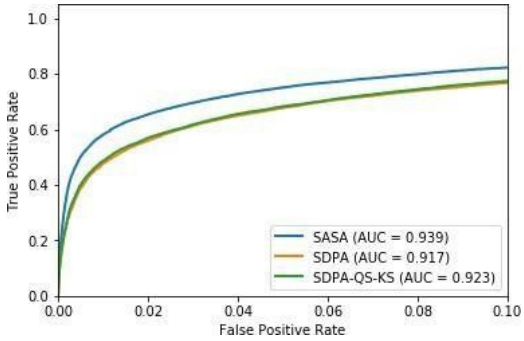


Fig. 5. ROC curves of models on 'Dataset B' - test set for FPR 0.10.

Results on Geo Routes Classification

A comparison of our results using different types of attention layers is presented in Table IV. For each dataset, we conducted multiple experiments using different data types; coordinates and countries as described in Sec. III, and differentAttention layers; Scaled Dot Product Attention (*SDPA*), *SDPA- QS-KS* and Source-Aware Self-Attention (*SASA*), as describedin Sec. V-A. Furthermore, we also conducted one experiment using the *SDPA* as is, by just replacing the keys (or queries) with the sources (denoted by *SDPA-Q-S*), i.e.,

Attention $(\mathbf{Q}, \sigma(\mathbf{S}), \mathbf{V})$, (9)

and another experiment by using the *SDPA* as is and justmultiplying it with the sources (denoted by *SDPA-Mul-S*), i.e.,

Attention $(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \circledcirc \mathbf{S}$. (10) As Table IV shows all three *SDPA-QS-KS*, *SDPA-Q-S*, and

SDPA-Mul-S led to performance degradation (even relativeto the regular *SDPA* without the use of sources). For each experiment in Table IV, after training our neural network over the corresponding training set, we evaluate our method over thetest set, which consists of 20% of the dataset and basedon the 'combined' labeling.

Given the imbalance of the datasets, an "always GREEN" classifier will achieve an accuracy of 99.57% and 94.89% on datasets A and B respectively, with 0 false

alarm. Not surprisingly, all the tested methods achieved similar results, with *SASA* showing (slightly for 'Dataset A') better accuracy and better false alarm rate: an accuracy of **99.24%** with a false alarm of **0.8%** on 'Dataset A', and an accuracy of 90.19%, with 9.5% FA on 'Dataset B', as presented in Table IV. All the methods achieve high Recall values, where *SASA* achieves the highest results with 99.52% on 'Dataset A', while the "always GREEN" classifier would get 0%. Namely, looking at all three parameters, we can see that learning was achieved despite the large imbalance. Notice that on 'Dataset A', which is less noisy (the 'M' sources comprises about 74% of the sources), the *SASA* layer achieves 0.1-0.2% improvement, while on 'DatasetB', which is noisier (the 'M' sources comprises less than 59% of the sources), the *SASA* layer achieves 1.6-1.9% improvement.

It can also be noticed that the use of countries achieves better accuracy than coordinates; this may be explained by the fact that borders are far from convex, and it is hard to learn cases when one country stretches into another. For example, the Vladivostok area can be easily mistaken to be part of China, and distinguishing between Singapore and Batam, ID that are only 30Km apart is hard since Indonesia has territories that are engulfing Singapore from almost any direction.

Figures 4 and 5 present ROC curves of all models on 'Dataset A' and 'Dataset B', respectively, using coordinates as entities and based on the 'combined' labeling method. It can be seen that both *SDPA*, *SASA*, and *SDPA-QS-KS* achieve great TPR results on 'Dataset A' even for low FPR below 0.5%. On 'Dataset B' it can be seen that *SASA* outperforms all other models, and for low FPR, achieve TPR values higher than the rest by about 10%. The AUC value for *SASA* (see Table IV) is 99.80 and 94.56, for datasets A and B, respectively.

The purpose of detecting deflected routes is to be able to react in case of an IP hijack attack. If a system creates too many false alarms (FAs), the load on the responsible team may be too high, and system credibility will be hurt. Thus, we would like to control the false alarm rate and achieve the highest possible detection rate. Table IV shows trade-off points for moderate FA of 1%, and low FA rate of 0.1%. For countries, which is the better option as we have already seen, *SASA* has a significantly better detection rate, over 80.81% detection rate, and a gap above 10% for 0.1% FA for 'Dataset A'. For 'DatasetB', 0.1% FA is not a possible working point since all methods detect less than a third of the deflection events; for 1% FA, *SASA* detects 63% of the deflections, about 3% better that *SPDA*.

Table V displays a comparison of experiments using different labeling methods: Geo, Owner, VF, and Combined as described in Sec. III. In each experiment, we trained our network with coordinate data using one of the specific labeling methods, and evaluate its performance based on a unified test set using the 'combined' labeling method. The results highlight the generalization ability of our method; by training our network using 'Geo' and 'Owner' labels, our method can achieve high accuracy for the 'combined' labeling method, in some cases even better results than by training using the 'combined' labeling method. It can also be seen that based on the VF labeling, our network achieves lower accuracy, which may be explained by the fact that VF is not based on geographical data.

Exploration of Source Scores

Table VI displays a comparison of source scores that were calculated by *SASA* for the different path types. Interestingly,

moved to Hong Kong. While we do not suspect this routing change is malicious due to the parties involved, this is exactly the type of misconfiguration a hijack alert system should flag. Figure 7(c) shows that *SPDA* managed to highlight HK as the problematic part of the route, while *SASA* flagged the US origin.

Bangladesh Leak

On September 26th, 2018, between about 7:45 UTC and 14:45, BTCL (AS17494) of Bangladesh leaked over 3500 APs to Telecom Italia (AS6762) that exported to its peers. As a result, many routes worldwide were diverted to Bangladesh.

The example in Figure 7 shows a route between the Total cloud in Los Angles, California, to a Colt IP address in Dublin, Ireland. The route before the leak was comprised of Total (AS46562) in LA, GTT in LA, Level 3 from LA to Paris, and Colt (AS8220) from Paris to Dublin.

Both the *SASA* and *SDPA-QA-KS* layers successfully highlighted the Bangladeshi hop as the deflection source. *SDPA* had the same score for the Bangladeshi and Irish portions of the route with a similar score for the Singaporean portion.

Asian Hijack

In 2016 China Telecom hijacked traffic from several European countries to an Asian government network (details are anonymized). In this example, we show the route from the GARR academic network, where Cogent carry the traffic from Rome, Italy to Los Angeles; there it is peering with China Telecom that hijacks the traffic through Guangdong, China.

Both the *SASA* and *SDPA-QA-KS* layers successfully classi- fied the route as hijacked. Furthermore, figure 7(e) shows that *SPDA* managed to highlight China as the problematic part of the route, with a relatively high attention score.

Geo-Deflection in Europe

On April 2017, a geographic analysis detected a deflection of a route towards a single AP that belongs to a tier-2/3 provider in New England from a large cloud provider in France. The AS- level route, both the one obtained from traceroute and from BGP announcements, was benign:

Hur- ricane Electric (AS6939) connects two customer networks. Therefore, an analysis based on the AS-level route (such as 'VF') would not flag this route. However, the geography of theroute, traversing Kiev, was highly unusual and extremely suspicious.

The route was compared to many other routes betweenthe French provider in France and other destinations in New England, all these routes where geographically confined to West Europe and North America. Following a message we sent to the French provider NOC, the route was immediately corrected.

As presented in Figure 8, *SASA* successfully classified the route as hijacked and highlighted Ukraine as the problematic part of the route, with a very high attention score.

Examples Summary
In general, both the *SASA* and *SDPA-QA-KS* layers were successful in highlighting the cause of the deflection in the routes, but *SASA* was **better at detecting deflection** events.

There is a subtle issue regarding the highlighting of the problematic portion of the route. Consider the American hop on the route from Sweden to Hungary (Figure 7(b)). Thereason theUS hop is problematic is that it appears in a continental route inEurope, and there is nothing wrong with its location on its own. In other words, what is problematicin the route is the triplet Sweden (source), US ⟨(middle), Hungary (destination).

In practice, this is wh⟩at we would like the network to learn. Remember, that nowhere in the training process we directed the network what is wrong, each route was simply labeled as 'GREEN' or 'RED'.

5. DISCUSSION

We showed that by introducing an attention mechanism to the model, deflected routes' detection rate improves. We also showed that in many cases, the attention mechanism highlights the problematic portion of the route successfully. However, as discussed in Sec. VII-H, the reason to flag a route as deflected is not a single segment in it, but the combination of this segment with source and destination location. As future work, it will be interesting to force the system to output a triplet, or alternatively, to disallow it to consider either end of the route as problematic.

Selecting the right false-alarm rate for the system is not trivial. If the route monitoring system generates too many false alarms, it will lose operators' credibility or simply overwhelm them with work. Assume that an organization wishes to protect 200 APs each from 50 locations, this results in 10,000 routes to monitor, and with FA = 0.01, it may result in 100 FAs. However, routes are highly correlated since routes towards a destination share the final portions of the route, and routes emanating from a monitoring point share the same start. In addition, organizations tend to have several APs in the same location connected to the same upstream providers. This will significantly reduce the number of events for the SOC team to handle, where an event is the collection of all flagged routes with the same routing problem at the same time. Of course, with time, a noisy dataset like 'Dataset B' will gradually be cleaned since false alarms will trigger database corrections.

We need to acknowledge that traceroute measurements come with their own problems. Some networks block ICMP probes [32, Sec. 4.3], but this mostly happens at stub ASes, which are less important for detecting deflections. Table II shows that only 4-6% (X data source) of the routers are either private IP addresses or do not return our probes. This number is quite small, and is mostly due to a few transit networks. SASA associates a weight to these types of routers, and it may be interesting in the future to separate this 'dataset' to non- responsive and private IPs.

Another problematic issue in this paper is the lack of external ground truth. We strongly believe that our labeling, at least for the 'solid' dataset, is mostly reliable, since this data is used and eye-balled continuously for monitoring routes. Indeed, the results for the solid dataset are significantly better than for the noisy dataset.

6. CONCLUSION

In the conclusion, you should summarize the main findings of your study, restate your research question and objectives, and provide recommendations for future research on the topic.

In order to geolocate the IP addresses, we used multiple geolocation services, with various levels of confidence; all suffer from geolocation errors. To take advantage of the knowledge of the sources, we developed an attention-based layer that aimed to deal with multi-source data; we termed it Source-Aware Self-Attention, *SASA*. This layer also highlights the cause of flagging out a problematic route.

We showed that both the *SASA* and *SDPA* layers were successful in highlighting the cause of the deflection in the routes, but *SASA* was better at detecting deflection events. We showed that by training our network on an unbalanced dataset,we could detect hijacked routes with an accuracyof 99.24% with 0.80% False Alarm and a detection rate of 99.52%, based only on geographical data. We also tested *SASA*and the other models on a few interesting deflections cases from 2017-2020, and correctly identified all of them as hijacked.

Finally, a sophisticated attacker may attempt to hide the attack by sending ICMP replies that impersonate a legitimate route. This makes the attack significantly harder to design and perform. However, an interesting future

direction is to extend the input SASA with delay and TTL in order to make such impersonation harder.

7. REFERENCES

[1] P. Sermpezis, V. Kotronis, A. Dainotti, and X. Dimitropoulos, "A survey among network operators on BGP prefix hijacking," ACM SIGCOMM Comput. Commun. Rev., vol. 48, no. 1, pp. 64–69, 2018.

[2] C. Demchak and Y. Shavitt, "China's maxim–leave no access point unexploited: The hidden story of China telecom's BGP hijacking," Mil. Cyber Affairs, vol. 3, no. 1, p. 7, Jun. 2018.

[3] M. Lad, D. Massey, D. Pei, Y. Wu, B. Zhang, and L. Zhang, "PHAS: A prefix hijack alert system," in Proc. USENIX Secur. Symp., 2006, vol. 1, no. 2, p. 3.

[4] J. Qiu, L. Gao, S. Ranjan, and A. Nucci, "Detecting bogus BGP route information: Going beyond prefix hijacking," in Proc. Secur. Privacy Commun. Netw., 2007, pp. 381–390.

[5] P. Sermpezis et al., "ARTEMIS: Neutralizing BGP hijacking withina minute," IEEE/ACM Trans. Netw., vol. 26, no. 6, pp. 2471–2486, Dec. 2018.

[6] S. Cho, R. Fontugne, K. Cho, A. Dainotti, and P. Gill, "BGP hijack- ing classification," in Proc. Netw. Traffic Meas. Anal. Conf. (TMA), Jun. 2019, pp. 25–32.

[7] A. Edmundson, R. Ensafi, N. Feamster, and J. Rexford, "Character- izing and avoiding routing detours through surveillance states," 2016, arXiv:1605.07685. [Online]. Available: http://arxiv.org/abs/1605.07685

[8] A. Shah, R. Fontugne, and C. Papadopoulos, "Towards characterizing international routing detours," in Proc. 12th Asian Internet Eng. Conf., New York, NY, USA, Nov. 2016, pp. 17–24.

[9] M. Syamkumar, R. Durairajan, and P. Barford, "Bigfoot: A geo-based visualization methodology for detecting BGP threats," in Proc. IEEE Symp. Visualizat. Cyber Secur. (VizSec), Oct. 2016, pp. 1–8.

[10] S. Papadopoulos, K. Moustakas, and D. Tzovaras, "BGPViewer: Using graph representations to explore BGP routing changes," in Proc. 18th Int. Conf. Digit. Signal Process. (DSP), Jul. 2013, pp. 1–6.

[11] G. Theodoridis, O. Tsigkas, and D. Tzovaras, "A novel unsupervised method for securing BGP against routing hijacks," in Computer and Information Sciences. London, U.K.: Springer, 2013, pp. 21–29.

[12] S. Papadopoulos, G. Theodoridis, and D. Tzovaras, "BGPfuse: Using visual feature fusion for the detection and attribution of BGP anomalies," in The 10th Workshop Vis. Cyber Secur., Oct. 2013, pp. 57–64.

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Systematic Review on AI-Blockchain Based E-Healthcare Records Management Systems

MRS. K. KAVITHA[1], AMUTHA P[2]
[1]Associate Professor, [2]Student
[1,2]Department of Compute Science & Engineering
Annai Mathammal Sheela Engineering College, Salem, Tamil Nadu, India

ABSTRACT

Electronic health records (EHRs) are digitally saved health records that provide information about a person's health. EHRs are generally shared among healthcare stakeholders, and thus are susceptible to power failures, data misuse, a lack of privacy, security, and an audit trail, among other problems. Blockchain, on the other hand, is a groundbreaking technology that provides a distributed and decentralized environment in which nodes in a list of networks can connect to each other without the need for a central authority. It has the potential to overcome the limits of EHR management and create a more secure, decentralized, and safer environment for exchanging EHR data.

Index Terms – Artificial Intelligence, E-Healthcare.

1. INTRODUCTION

Medical and healthcare researchers emphasize the importance of their ability to collect and analyze multi-source data in order to identify potential community health hazards, provide case-specific therapies, and deliver focused medicine, which could promote informed clinical decision making and lead to improved patient care quality. This information can help to improve personal health information systems such as patient health records (PHR) and patient portals. Patients frequently do not have easy access to their historical data, while clinicians retain primary ownership.

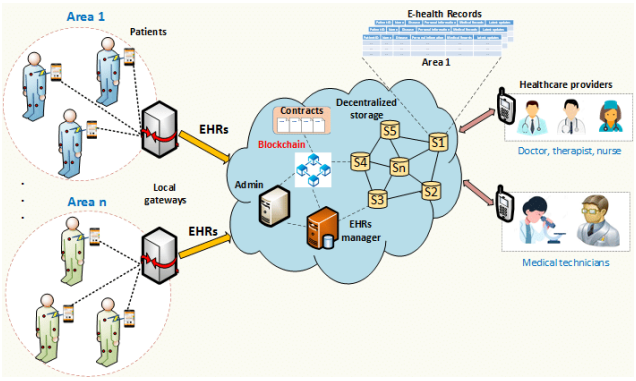


Fig. 1: Work Flow

Healthcare systems used networked electronic health management practices combined with clinical imaging systems to help doctors get more reliable, accurate, and timely access to patient's data.

2. EXISTING SYSTEM

We can now use the blockchain framework or existing platforms to develop decentralized applications. The most popular are Hyperledger, which both allow developers to construct new blockchain applications on top of current ones and request that they create new test nets using the protocol. The use of blockchain has evolved rapidly in recent years. A system that was created for cryptocurrencies is now being used to cast votes due to its temper-proof characteristics. The use of smart contracts enforces accountability for all parties involved and ensures the contract's integrity. The data security issue of IoT systems has been solved through the utilization of blockchain-based data communication and storage systems. The limit to the applicability of blockchain is the imagination of the user. While in this study, our focus is to review the impact of blockchain in the healthcare sector, there is no difference between the application of blockchain in healthcare and the general sector as the healthcare sector uses all the services offered by blockchain, such as money transfer, personal data security, logistics, and overall data safety.

3. DISADVANTAGE

To tackle health information records and exchange, a system is required to be developed, managed, and maintained. A third party develops and maintains traditional personal health record and electronic health record systems, with trust, privacy, and data security remaining important challenges. However, the third-party-based existing healthcare recording systems cannot satisfy stakeholders' privacy needs. As a result, the traditional electronic healthcare model lacks transparency because of privacy and data security issues.

4. PROPOSED SYSTEM

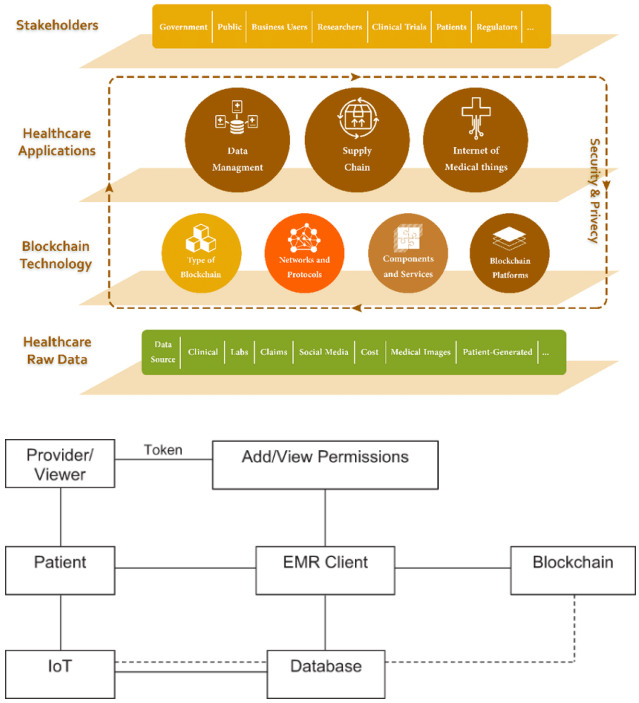
An architectural design that contains two layers is proposed: (i) one is flow and storage controls of the data; (ii) the second one is a central healthcare data unit. The proposed methods work with the patient's own sensor network, where generated data are transferred to the healthcare system via a smart device. The generated data are transferred to the server, which acts as an agent for the

patient. It also works for data control, mining, and security management on the system. The authors of discuss the test and measure of architectural design in a variety of contexts, including man-in-the-middle assaults, denial-of-service attacks, and the implications for patient privacy. Prior to these experiments, several industries used various types of selection algorithms.

5. ADVANTAGES

Many parameters had to be determined before the testing could be carried out, beginning with the mining and mining selection algorithms. This analysis showed that processors are used at 25% and memory needs are at 95 MB, but before that, the network needed three miners. On the other hand, safety test subjects were selected based on the aforementioned assault and were likened The network's physical characteristics were also examined, such as processing time, overhead, and throughput in kbps (kilobytes per second). In this test, it was determined that processing and overhead costs were lower than in previous tests such as baseline.

6. DATAFLOW DIAGRAM



7. CONCLUSION

This research aimed to conduct a thorough review, survey, and categorization of relevant research papers on blockchain and their integration into various healthcare applications where certain literary patterns may be detected. This paper presented the bibliometric and functional distribution of 144 research papers on blockchain in healthcare. We evaluated the distribution of blockchain platforms and the various kinds of blockchain techniques used or proposed in the examined papers. The blockchain platform allows the development of decentralized applications where the pattern of data transfers is uncontrollable by any third-party organization. The data transactions of the entities are kept in a decentralized database in a verifiable, secure, immutable, and transparent way, along with a timestamp and other pertinent information. Additionally, blockchain technology has a variety of potential applications in healthcare, including data sharing, log management, medication, biomedical research and teaching, remote patient monitoring, and health data analytics. Even though blockchain adds many valuable features to healthcare applications, it has some drawbacks. We also analyzed the proposed solution in the reviewed papers for these drawbacks. Despite the considerable interest in blockchain technology, we discovered that its effect on healthcare applications is mostly in the documentation phase. There is yet to be a significant amount of study conducted in this area, as well as healthcare applications built on blockchain.

8. REFERENCE

[1] McClean, S.; Gillespie, J.; Garg, L.; Barton, M.; Scotney, B.; Kullerton, K. Using phase-type models to cost stroke patient care across health, social and community services. Eur. J. Oper. Res. 2014, 236, 190–199. [CrossRef]

[2] Soltanisehat, L.; Alizadeh, R.; Hao, H.; Choo, K.K.R. Technical, Temporal, and Spatial Research Challenges and Opportunities in Blockchain-Based Healthcare: A Systematic Literature Review. IEEE Trans. Eng. Manag. 2023, 70, 353–368. [CrossRef]

[3] Xing, W.; Bei, Y. Medical Health Big Data Classification Based on KNN Classification Algorithm. IEEE Access 2020, 8, 28808–28819. [CrossRef]

[4] Khan, A.A.; Wagan, A.A.; Laghari, A.A.; Gilal, A.R.; Aziz, I.A.; Talpur, B.A. BIoMT: A State-of-the-Art Consortium Serverless Network Architecture for Healthcare System Using Blockchain Smart Contracts. IEEE Access 2022, 10, 78887–78898. [CrossRef]

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Pollution Control System & Public Safety Protection Using IoT and Big Data Privacy

DR. G. KARTHIK[1], MRS. T. GEETHA[2], C. SIVAKUMAR[3]
[1]Professor, [2,3]Assistant Professor
[1]Department of Information Technology, [2,3]Department of Compute Science & Engineering
[1]Karpagam College of Engineering, Coimbatore, Tamil Nadu, India
[2,3]VMKVEC, Salem, Tamil Nadu, India

ABSTRACT

To control the pollution and manage public safety protection from pollution through IoT and Big data. The Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Humankind, moving to a period centered upon improvement has overlooked the significance of supportability and has been the real guilty party behind the rising pollution levels in the world's air among all other living life forms. The pollution categorized into air pollution, water pollution and noise pollution. The pollution levels at certain spots have come to such high degrees that they have begun hurting our very own. An IoT based pollution observing framework incorporates a sensor interfaced to a outfitted with an WLAN connector to send the sensor perusing to a Thing Speak cloud. Further extent of this work incorporates an appropriate AI model to foresee the pollution level and an anticipating model, which is fundamentally a subset of prescient displaying. As age of poisonous gases from ventures, vehicles and different sources is immensely expanding step by step, it winds up hard to control the dangerous gases from dirtying the unadulterated air, water related pollution & land related pollution. In this system a practical pollution observing framework is proposed. This framework can be utilized for observing pollutions in conduct of specific territory and to discover the pollution peculiarity or property examination. The obligated framework will concentrate on the checking of pollution poisons concentrate with the assistance of mix of Internet of things with wireless sensor systems. The investigation of pollution quality should be possible by figuring Air Quality Index (AQI), Water Quality Index (WQI) and Land Quality Index (LQI). This system attempts to save the natural resources available for public safety protection kind by continuously checking the quality air, monitoring the status of the soil, the pollution can be controlled and thereby increase the public safety. Also, by knowing the air, water, land moisture and temperature volume of contents are maintained through big data.

Index Terms – Pollution Control System, IoT.

1. INTRODUCTION

Pollution can be characterized as nearness of moment particulars that bothers the working of common procedures and furthermore creates unfortunate wellbeing impacts. In another way contamination can influence the characteristic periodicity and furthermore can irritate the wellbeing of person. As modernization and automation is becoming in all respects widely Pollution is likewise getting presented everywhere way.

It has been seen that in mechanically creating or created nations human wellbeing get significantly influenced due to Air Pollution and Water pollution where there is no framework to screen it or monitor it.

In late explores it has been demonstrated that there is a high connection batten's climatic toxins and maladies like asthma and lung related ailments. Air Pollution and Water pollution is currently a noteworthy worry over the globe and WHO has built up specific rules to confine the cutoff points of specific gases like O3, NO2, SO2.

The Air Quality Index and Water Quality Index estimation and Pollution observing are mostly done surface stations that are essentially exact and precise. They show ideal unwavering quality and viable in estimating a wide scope of air toxins. Be that as it may, even after every one of these stations slack fundamentally in three territories:

- Infrastructure, essential for establishment as a result of the colossal size
- Operational necessities are basically mind boggling
- The common costs of setting up, day by day support and alignment

The terms monitoring and assessment are frequently confused and used synonymously. The process of industrial quality assessment is an evaluation of the industrial quality in relation to standard quality set by pollution control board. Due to the complexity of factors determining industrial quality, large variations are found between different industries.

Similarly, the response to industrial impacts is also highly variable. To design an Industrial machine control and monitoring system using IOT. Surveillance is most important security systems in home, industrial, office and public places. To build a robust system that can measure the industrial pollution and help to reduce it and to decrease human interference in monitoring the industrial pollution to reduce pollution and provide a healthy environment for the workers to work in. To build a robust system that evaluates the industrial pollution continuously and indicates when there is an increase in emission and controls it using IOT.

This system has the following modules:

- Client Control Module
- Server Control Module
- Air Measurement Module
- Water Measurement Module
- Land /Soil Measurement Module
- IoT'S Connected Module
- Public Safety Protection Module
- Progress Report Control Module

Through this system administrator can see the types of pollution details prevention of public safety protection problems previously.

2. EXISTING SYSTEM

The Existing system has known internet are inter-connected devices to the internet. To fulfill the need of flourishing monitoring system, in our project is establishing a network called Internet of Things, in which sensing devices are connected.

Air pollution is not only natural medical matters impact on creating nations alike. The strong effect of air pollution on wellbeing are extremely mind blowing as there are a broad area of sources and their particular influence differ from one another. The synthetic substances reason an assortment of mankind and natural medical issues enlarge in air contamination impacts on condition also on human wellbeing. To screen this contamination WSN framework is expressed.

The proposed framework comprises of a Unit of Mobile-DAQ and a fixed Internet-Enabled contamination observation System. The Mobile-DAQ unit incorporates a solitary chip microcontroller, air pollution sensors exhibit, and GPS Device. The Pollution-Server is a top of the line individual computer application server with Internet network. The Mobile-DAQ unit assembles air toxins levels (CO, NO2, andSO2), and packs them in a casing with the GPS physic distribution, time, and date. The reason is to send the Pollution-Server by means of zig bee device. The pivotal-Server is interacting to Google Maps to show the area of equipment. It can associate database server to the Pollution-Server for putting away the toxins range for future utilization by different user, for example, condition security offices, vehicles registration experts, and vacationer and insurance agencies.

SIM900A is an ultra-compact and reliable wireless module. The SIM900A is a complete Dual-band GSM/GPRS solution in a SMT module which can be embedded in the customer applications. Featuring an industry-standard interface, the SIM900A delivers GSM/GPRS 900/1800MHz performance for voice, SMS, Data, and Fax in a small form factor and with low power consumption. With a tiny configuration of 24mmx24mmx3mm, SIM900A can fit in almost all the space requirements in user applications, especially for slim and compact demand of design.

With wireless embedded computing system. Internet of Things is a technology that hook up the sensors with embedded system and allow the data from these sensors to travel over an Internet. We are implementing developing model which is able to monitors the inconstancy of parameter like Air, Noise, Temperature, Humidity and Light. In the proposed model we use microcontroller ATMEGA328 that is mounted on Arduino Uno board. We are using 5 sensors, MQ-7 as a gas sensor. It detects the concentration of carbon monoxide in air. To measure the fluctuations in noise levels we use M213 high sensitivity microphone sensor module. LM35 is used as a temperature sensor and SY- HS220 as humidity sensor. To measure the intensity of light LDR sensor is used. The values are displayed in the LCD display.

Drawbacks of Existing System

- No devices are connected
- Maintenance expenses are high
- Not Expertise
- No Public Safety
- Slow process
- Highly watchable
- No proper measurement reading for air and water

3. PROPOSED SYSTEM

The proposed system focuses IoT for the most part manages associating shrewd gadgets (implanted hardware gadgets) to Itb by tackling the upside of OSI layered Architecture. With regards to this work. It proposes a group of Air Quality and Water Quality Monitoring Sensor bits, which are utilized to quantify the convergence of Air contaminations noticeable all around. All the Air Sensors are interfaces with a minor implanted stage outfitted with system availability and are interconnected making it a worldwide system of associated things.

The proposed system aims in designing a robust system that monitors real time emission levels and temperature of all the industries and required areas, store all the collected data in and analyze them in cloud using Internet of Things. Our system uses various sensors such as temperature sensor, DHT sensor, MQ-2 sensor, MQ-6 sensor, Dust sensor to measure various parameters such as temperature, gas, dust respectively.PIC microcontroller is used which uses Reduced Instruction Set Computing (RISC) program that reduces the complexity.

DHT22 Sensor
This DHT22 is a temperature & a humidity sensor with a digital signal output. It provides high stability and reliability. It consists of a Negative temperature coefficient temperature measuring component and a resistive type humidity measurement component. It can be connected to a microcontroller and offers quick, anti-interference ability and cost-effectiveness.

MQ-2 Sensor
A carbon monoxide analyzer or CO analyzer is a device that detects the presence of the carbon monoxide gas in order to prevent carbon monoxide poisoning. The circuit setup consists of analyzer head connected to an amplifying unit. A number of supporting resistances are used to avoid voltage drop across the circuit. Resistance value of MQ-2 is difference to various kinds and various concentration gases. So, when using these components, sensitivity adjustment is necessary.

It is recommended that calibrating the detector for 200ppm CO in air and using Load resistance of about 10KΩ (5KΩ to 47 KΩ) increases circuit efficiency.

MQ-6 Sensor
The MQ-6 Sensor can detect the small particles like cigarette smoke and it can distinguish small particles like smoke from large house dust by pulse pattern of signal output. It is used for detection of dust in the air, indoor air quality monitoring. The features are compact size and light weight (about W59x H46x D18 mm, 20g). It works on the principle of PWM (pulse width modulation) output (Low pulse output).

Liquid Crystal Display (LCD)
A liquid-crystal display (LCD) is a flat panel display or other electronically modulated optical device that uses the light modulating properties of liquid crystals. Liquid crystals do of the code significantly. WIFI module ESP8266 is used to store the data in the cloud which is flexible and easy to connect and it is connected through the hotspot. The data can be viewed in any browser including smart phones by logging in using the credentials.

4. ADVANTAGES OF PROPOSED SYSTEM

- Sensors and Programming Quality
- The system is interconnected with Internet (i.e) IoT's
- The reading data should be stored in "Big Data"
- Safety no defects
- Savings
- Time Consuming
- Producing the reports one department progress completion.

The proposed system is designed to overcome all the disadvantages of the existing system.

5. MODULES

"Pollution Control System & Public Safety Protection Using Iot & Big Data Privacy" has the following modules:

- Client Control Module
- Server Control Module
- Air Measurement Module
- Water Measurement Module
- Land /Soil Measurement Module
- IoT'S Connected Module
- Public Safety Protection Module
- Progress Report Control Module

6. CLIENT CONTROL MODULE

Client control module is intended to cleanup and "hide" options from clients that could potentially break a site after site handoff. The first iteration/release of this module simply provides a settings form to select certain themes which should be hidden from view on the theme listing page. This module is inter connected with the IoT and server control module.

7. SERVER CONTROL MODULE

Server control module can control the server upto three positions with push button switches or toggle switches. This can be used for a variety of different applications such as:

- Opening level crossing barriers
- Opening goods shed doors
- Opening gates
- Switching points

AIR MEASUREMENT MODULE
This module is used to air measurement module comprises an antenna, adapted to receive a first measuring signal from a device under test or adapted to transmit a second measuring signal to the device under test. The readings based on the Air Quality Index measuring values.

WATER MEASUREMENT MODULE
This module is used to water level sensor module. This water level sensor module has a series of parallel exposed traces to measure droplets/water volume in order to determine the water level. Very Easy to monitor water level as the output to an analog signal is directly proportional to the water level. The readings based on the Water Quality Index measuring values.

LAND /SOIL MEASUREMENT MODULE

This module is used to measure the volumetric water content in soil. Since the direct gravimetric measurement of free-soil moisture requires removing, drying, and weighing of a sample, soil moisture sensors measure the volumetric water content indirectly by using some other property of the soil, such as electrical resistance, dielectric constant, or interaction with neutrons, as a proxy for the moisture content. The readings based on the land or soil measuring values.

PUBLIC SAFETY PROTECTION MODULE
This module is used to create the safety to public from the

dangerous. The main aim is to protection of public and give the guidelines to handle the hazards situation.

PROGRESS REPORT CONTROL MODULE

This module is used to create the modules and sub module reports. The progress report is to control the time series.

8. IMPLEMENTATION

Implementation is the stage, which is crucial in the life cycle of the new system designed. **"A CLOUD ASSISTED ZIGBEE-BASED ZOO-ANIMAL HEALTH MONITORING SYSTEM USING BIG DATA AND IOT SERVICES"** is used to implement the experimental validation of animal health monitoring system (AHMS) which is capable to the measuring of body temperature, rumination, and heart rate parameters with environmental parameters (surrounding temperature and humidity). The system is based on the IEEE 1451.2, IEEE 802.15.4, and IEEE1451.1 standards. The PIC18F4550 microcontroller and XBee-PRO S2 module were used to the development of AHM system. The four-sensor module such as body temperature, heart rate, surrounding humidity and temperature and rumination has been successfully developed.

9. CONCLUSION

The Arduino as a motherboard is chosen as the processor. Temperature sensors, smoke sensors, and radiation sensors are the sensors connected to the processor. The smoke sensor senses the smoke if it is greater than 613 it informs the processor and the processor immediately sends a message through GSM with the temperature recorded at that time. The same process will be done if the radiation sensor senses the radiation above the range 250, whether small scale or large scale should and must have this system to monitor the emissions.

10. REFERENCES

[1] POLLUTION CONTROL TECHNOLOGIES – Vol. I - Pollution Control Technologies - B. Nath and G. St. Cholakov
©Encyclopedia of Life Support Systems (EOLSS) review on engines, fuels, illustrated with numerous case studies].

[2] Glassman, I. (1996), Combustion, 3rd ed., Academic Press, Inc., London, UK. [Thermodynamic and chemical fundamentals of combustion].

[3] Handbook of Air Pollution from Internal Combustion Engines (1998), Ed. E. Sher, 663 pp. San Diego, CA, USA: Academic Press.
Heck R. M. and Farrauto R. J. (2002). Catalytic Air Pollution Control: Commercial Technology. 416 pp. New York, USA: John Wiley & Sons [Catalysts for pollution control].

[4] http://www.epa.gov/ and http://eea.eu.int/ [the websites of the US and the European agencies with abundance of information and links] http://www.epa.gov/oeca/sector/ [profiles and reviews of the industries covered in the Theme]

[5] http://www.wikipedia.org/wiki/ [A comprehensive encyclopedia explaining numerous themes of physics, chemistry and other sciences]

[6] http://www.wiley-vch.de/vch/software/ullmann/. Ullmann's Encyclopedia of Industrial Chemistry – 7th edition.

[7] Kalhammer, F. R., Prokopius, P. R., Roan, V. P.. and Voecks, G. E. (1998). Status and Prospects of Fuel Cells as Automobile Engines. A Report of the Fuel Cell Technical Advisory Panel, Section II, 19 pp., Prepared For State of California Air Resources Board, Sacramento, California, USA.

❑❑❑

e-ISSN: 0976-7967, p-ISSN: 2319-2216 © VSRD International Journals: www.vsrdjournals.com

Survey On Solanum Lycopersicum Fronds Malady Detection and Pesticides on Android App

RABINTHA J
Department of Compute Science & Engineering
Mahendra College of Engineering, Salem, Tamil Nadu, India

ABSTRACT

Tomato is the most popular crop in the world, it is found in different forms irrespective of the cuisine. After potato and sweet potato, it is the crop which is cultivated world widely. India ranked second place in the production of tomato. However, the quality and quantity of tomato crop is decreasing day by day due to different diseases which affect the crop. This meets the farmer with heavy losses. To decrease this loss, it is very much necessary to have a complete supervision over the growth of the crop. There are various categories of diseases that harm this tomato leaves on a very large scale like Late Blight, Bacterial Spot, Early Blight, Septoria Spot, Mosaic Virus etc. So, it is necessary to get a solution to prevent these diseases before affecting the crops. This could be done only if we can detect the disease when it is in its beginning stage. With the help of different deep learning algorithms, leaf disease detection can be done very easily and more precisely with great results and with very accurate pesticides. In this paper, Convolutional Neural Networks (CNN) is applied to the dataset. The CNN shows the best accuracy than other algorithms. And these models will deploy in the Android app for users/farmers convenient and easy purpose

Index Terms - Tomato Leaf Disease, Deep Learning, CNN, Detection, Pesticides, Android App.

1. INTRODUCTION

Plant diseases are a major cause of crop destruction that is detrimental to the economy. Plant diseases harm not only humans, but also herbivores. It is estimated that about 40% of crops are lost due to plant diseases. In agricultural countries, where most of the population depends on the agricultural sector, plant disease detection plays an essential role in boosting crop yields and boosting the economy. According to a report by the United Nations Food and Agriculture Organization (FAO), the world's population is estimated to grow to about 9 billion people in a few years. However, production within the agricultural sector needs to be increased by at least 70% to meet human food needs [1].

Image processing can be used for automatic detection of various diseases. Image processing plays an important role. It is used for plant disease detection as it provides the best results and reduces human effort. Image processing is the process of converting an image into pixel form or binary number form and then performing some operations to extract useful information from the image for further processing. A method of performing some machinations on an image in order to obtain an enhanced image or to extract useful information from it. It is a type of signal processing where the image is a input and the output is the properties/features of the image associated with that image [3].

2. LITERATURE REVIEW

In 2019 Amrita S. Tulshan, together with Nataasha Raul, published a paper here addressing the detection of plant diseases. They applied the K Nearest Neighbor classification (KNN) is a machine learning algorithm and obtained a good accuracy of 98.56% in predicting plant leaf disease [4]. Arti N. Rathod, Bhavesh Tanawal, and Vatsal Shah published a research paper on leaf disease detection in 2013, describing different methods to detect affected leaves using image processing techniques [5]. Prajwala TM, Ala

Pranathi, Kandiraju Sai Ashrita, Nagaratna B. Chittaragi, Shasidhar G. Koolagdi, 2018

Having published a paper in which they were studying tomato leaves for disease detection, they

A CNN model known as Le Net was slightly modified to find and classify tomato leaf diseases with an average accuracy of 94–95% [6]. In 2020, Surampalli Ashok, et.al published a paper on the detection of tomato leaf diseases. Accuracy of 92.94% and 98.12% respectively [7]. Mohit Agarwal, et.al in 2020, Gupta published a research paper on the detection of tomato leaf diseases using CNNs, obtaining an accuracy of 91.2% [8].

Halil Durmuu, et.al published a paper on tomato leaf disease detection in which they applied Alex Net and Squeeze Net with accuracies of 95.65% and 94.3%, respectively [9]. Konstantinos P. Ferentinos published a paper in 2018 addressing plant leaf disease detection and applied the VGG model to achieve 99.48% accuracy [10]. In 2017, Alvaro Fuentes, et.al published a research paper on "A Robust Deep Learning-Based Detector for Real-Time Tomato Plant Disease".

Here we apply VGG 16, resulting in an accuracy of 83.06% [11]. Geetharamani G. and Arun Pandian J. published a paper in 2019.

A 9-layer deep CNN was used to detect plant leaf diseases, achieving 96.46% accuracy [12]. In 2019, Peng Jiang Yuehan Chen, Bin Liu, Dongjianhe, Chunquan Liang

Apple leaf disease using a deep learning approach. of 91.2% [8].

Based on Improved Convolutional Neural Networks" here they were using VGG-FCN-VD16 and VGGFCN-S with average recognition accuracy 97.95% & 95.12%, respectively [13]. Xihai Zhangi et.al in the year 2017 presented a paper on "Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks", here they applied Google net model and got an accuracy of 98.9% [14]. Melike Sardogan, et.al presented a research paper in the year 2018 on "Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm" and got an average accuracy of 86%[15].

Utkarsha N. Fulari, Rajveer K. Shastri, Anuj N. Fulari presented a paper on "Leaf Disease Detection Using Machine Learning"in the year 2020, here they applied CNN model on grape dataset and got an accuracy of 99.7% and when applied onstrawberry dataset got an accuracyof 100%[16].

3. ALGORITHM

Convolutional Neural Networks (CNN) is a deep learning algorithm. A Convolutional Neural Network has three layers. That layers are a convolutional layer, a pooling layer, and a fully connected layer. Convolutional layer: produces an activation map by scanning the pictures several pixels at a time using a filter. Pooling layer: reduces the amount of data created by the convolutional layer so that it is stored more efficiently. Fully connected input layer – The preceding layers' output is "flattened" and turned into a single vector which is used as an input for the next stage. The first fully connected layer – adds weights to the inputs from the feature analysis to anticipate the proper label. Fully connected output layer – offers the probability for each label at the end.

In *Nishant* Shelar et.al, Fig A.1 CNN ARCHITECTURE is dicussed in [21].

VGG19 is a sophisticated CNN with pre-trained layers and a thorough grasp of how an image is defined in terms of form, color, and structure. VGG19 is a deep neural network that has been trained on millions of photos with challenging classification problems.

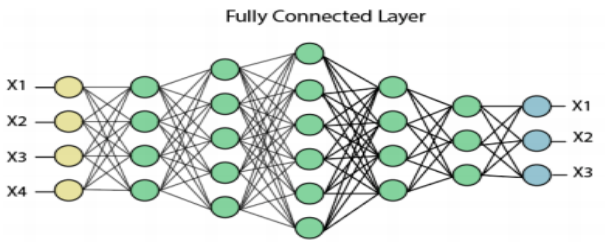


Fig.A.2. Fullyconnected layer

In *Nishant* Shelar et.al, Fig A.2 Fullyconnected layer is dicussed in [21].

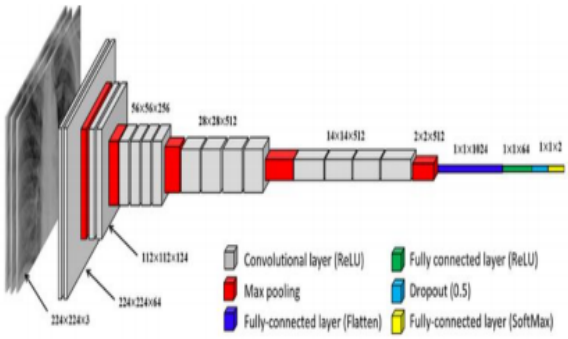


Fig.A.3. VGG-19

In *Nishant* Shelar et.al, Fig A.3 VGG-19 is dicussed in [21].

A 95.6% accuracy rate was achieved using early stopping while Training the model on 50 epochs. Figure A.4 depicts the visualization of training and validation accuracy. The result of detecting and recognizing a strawberry plant is shown in Figure A. 5. On the left, a healthy plant leaf, and on the right, a sick infected plant. The result of detecting and recognizing a potato plant is shown in Figure A.6. On the left, a healthy plant leaf, and on the right, a sick infected plant.

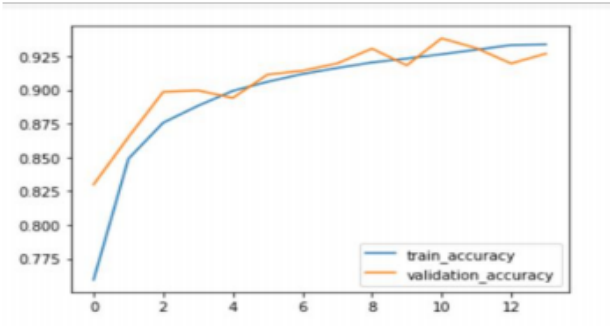


Fig.A.4.Training vs validation

In *Nishant* Shelar et.al, Fig A.4 Training vs validation

Tomato Disease	Trained Data	Tested Data
Bacterial spot	1702	425
Early blight	1921	481
Late blight	1852	464
Leaves Mold	1881	471
Mosaic virus	1791	449
Yellow Leaf Curl Virus	1962	491
Septoria leaf spot	1744	436
Target spot	1827	457
Spider mites	1741	435
Healthy	1926	481
Total	18,345	4585

Table A.1.The details of tomato diseases in the dataset

[17] consisting of tested dataset and trained dataset.

In Hsing-Chung Chen et.al, Table A.1.The details of tomato diseases in the dataset [17] consisting of tested dataset and trained dataset is dicussed in [22].

4. FUTURE ENHANCEMENT

In the beginning, data is collected from a website named "Kaggle" in the raw form [18]. After the pre-processing of the data, the required features are extracted as per the need, and then the data splits for training and testing purposes. After data pre-processing, data extraction process takes place. Training the data in the convolution neural network algorithm predicts very accurately the disease of leaves. By visualizing the image of leaf, it detects whether the leaf is caused by disease or not.

If the leaf is caused by the disease, CNN algorithm detects what disease it is. Then CNN also used to provide pesticides for the particular leaf disease by detecting process. The pesticides database is collected from the analysis or based on the historical disease event. By collecting the database of disease and its pesticides we can deploy these models on the smart technology as android app.

5. CONCLUSION

We have successfully created a disease classification technique that is used to detect leaf diseases in plants. A deep learning model that can be used for automatic detection and classification of plant leaf diseases is created.

Timely detection and identification of diseases which control the affect of leaf from diseases.

This paper presents a model where images have been pre-processed and then applied to deep learning algorithm Table A.2.Some parts of the test results were outputted from the built application by using the valid dataset [19]

6. REFERENCES

[1] Tomato Leaf Disease Detection Using Convolution Neural Network by Hareem kibriya, Rimsha Rafique, Wakeel Ahmad, S.M Adnan.

[2] Tomato Leaf Disease Detection Using Machine LearningbyRakesh Sharma, Ankita Panigrahi, Mamata Garanayak, Sujata Chakravarty, Bijay K. Paikaray and Lalmohan Pattanaik.

[3] Tomato Plant Leaf Disease Detection Using Convolutional Neural Network Radhika Dakhore, Shahjadi Sheikh, Aishwarya Masar, Aniket Zade, Prof. Aditya Bakshi.

[4] Tulshan, Amrita S., and Nataasha Raul. "Plant leaf disease detection using machine learning." 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2019.

[5] Rathod, Arti N., Bhavesh Tanawal, and Vatsal Shah. "Image processing techniques for detection of leaf disease." International Journal of Advanced Research in Computer Science and Software Engineering 3.11 (2013).

❑❑❑

A Case Studyon Metar Data Forecasting Using Time Series

GOKULAPRIYA V. ANUPPUR
Department of Compute Science & Engineering
Mahendra College of Engineering, Salem, Tamil Nadu, India

ABSTRACT

Forecast is a data science task that helps administration with capacity planning, goal setting and incongruity detection. Here Time Series algorithm is used to forecast the metar data also referred to as time-stamped data. These data points typically consist of successive measurements made from the same source over a fixed time interval and are used to track change over time. In addition, other time series models like ARIMAX, SARIMAX have many rigid data requirements like stationarity and equally spaced values. work with these requirements is highly difficult[1]. So, one of the time series algorithm ***Facebook Prophet*** *which giving quick, powerful, and accessible time-series modelling to data analysts and data scientist is discussed in reference [1]. The model can use it in the python although it can also be implemented in R.*

Index Terms - Metardataforecast, Faceprophet, Python, R..

1. INTRODUCTION

METAR stands for Meteorological Aerodrome Report. Visibility is an salient feature in all phases of flight, but especially when the aircraft is manoeuvring on or close to the ground. impoverished visibility at a destination can reduce capacity of airports leading to ground delays, flight diversions, flight cancellations and extra operating costs.

While analysing the data it takes much longer time to predict. In that case we require a quick and easy method to forecast the weather by having the metar data given by the data scientist or analyst.

Facebook prophet is the quick libraries developed by Facebook's Core Data Science team and the Facebook prophet is available in open source and is developed to forecast the time series data.

We can predict the result using the model for the upcoming hours that using independent features such as temperature, relative humidity, wind direction, wind intensity and so on.

2. LITERATURE REVIEW

Meteorologists forecast the weather and climate.

The meteorologist's study, analyse the metar data which is given by the predictor and they interpret, and provide weather reports and also a weather patterns on a day-to-day basis.

Atmospheric science trade in with the Earth's atmospheric conditions and its phenomena, such as the precipitation of typhoons, and snow. It also deals with anemometer variations of temperature, moisture, wind speed, direction, and similar patterns that produce distinct weather conditions.

Meteorologists predict weather using intricatedequations to manipulate the data.

They use anemometer to measure the speed and pressure of the wind. Instruments also include barometers (one of the most critical instruments in weather forecasting), rain gauges to measure rainfall, wind vanes to measure wind speed, thermometers to measure temperature, weather balloons, and weather satellites to compile the data required for forecasting.

At present, meteorologists prefer to use other tools to forecast the weather which include:[6]

- Doppler radar
- Satellite Data
- Radiosondes
- Automatedsurface-observing Systems
- Supercomputer
- Supercomputers
- AWIPS

These methods are referred from the reference link given in [6].

Here, the prediction can be made using the past data of weather that would be forecast using the time series algorithm that can be efficient and make the meteorologist know the earth's atmospheric phenomena.

There are several interesting facts are mentioned in this paper as follows, sales forecasting based on real-world data were analysed by Emir Zunic et al. [2] using Facebook prophet and 25 years of time series forecast using traditional algorithm called ARIMA (Autoregressive Integrated Moving Averages) by Jan G De Gooijer et al. [3], and forecasting international quarterly tourist flows using error-correction and time-series models by N. kulendran et al.[4] that include error-correction model into Australia from the major tourist markets of USA, Japan, UK and New Zealand.

Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India were analysed by Priya

Narayanan et al. [5] they analysis a period of 60 years of monthly Rainfall data for March, April, May (MAM) for six stations (Abu (Ab), Ahmedabad (Ah), Ajmer (Aj), Amritsar (Am), Bikaner (Bk), Jodhpur (Jd). The magnitude and practical significance of trend has been estimated using the Theil and Sen's median slope estimator, and assessing the percentage change over the mean for the period concerned (Yue and Hashino,2003; Yue et al.,2002.

3. ALGORITHM

Time series analysis comprises methods for analysing time-series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

Facebook Prophet generating time-series models that uses a few old ideas with some new twistsAt its core is the sum of three functions of time plus an error term: growth g(t), seasonality s(t), holidays h(t) , and error et :[1]

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

The growth function has three main options:

- **Linear Growth:** This is the default setting for Prophet. It uses a set of piecewise linear equations with differing slopes between change points. When linear growth is used, the growth term will look similar to the classic y = mx + b from middle school, except the slope(m) and offset(b) are variable and will change value at each changepoint.et [1]
- **Logistic Growth:** This setting is useful when your time series has a cap or a floor in which the values you are modelling becomes saturated and can't surpass a maximum or minimum value (think carrying capacity). When logisticgrowth is used, the growth term will look similar to a typical equation for alogistic curve (see below), except it the carrying capacity (C) will vary as afunction of time and the growth rate (k) and the offset(m) are variable and willchange value at each change point et.[1].
- **Flat:** Lastly, you can choose a flat trend when there is no growth over time (but there still may be seasonality). If set to flat the growth function will be a constant value.et [1]

For example:

$$g(t) = \frac{C(t)}{1 + x^{-k(t-m)}}$$

Consider a Metar dataset that includes Station, Timestamp, Air temperature, Dew point temperature, Relative humidity, Wind, direction, Wind speed, One hour precipitation, Pressure altimeter, Sea level pressure, Visibility, Wind gust, and so on

Installing library fbprophet and we extracting the required columns. Then constructing heat map that used toshow relationship between two variables.

The univariate time series has only one variable, a multivariate has more than two variables. After that as usualtraining and testing process takes place. Then, calculating the accuracy of tested data which are in the continuous data format.

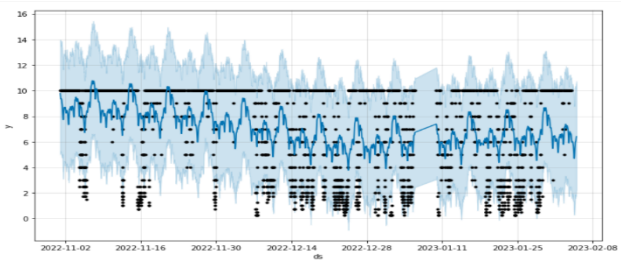


Fig. A.1. Visualizing in a heat map to know the relationship between two variables

The prophet procedure is an additive regression model with four main components:

A piecewise linear or logistic growth curve trend. Theprophet will automaticallydetect changes in trends by selecting changepoints from the data. It includes a yearly seasonal component modelled by using Fourier series and a weekly seasonal component modelled by using dummy variables.

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data.

It takes dependent and independent variables to plot and to know the yearly, weekly and daily seasonality with the help of Facebook prophet. In the algorithm ARIMA the same can be done but requires separate library. So, we going with the Facebook prophet to forecast the weather by the meteorologist.

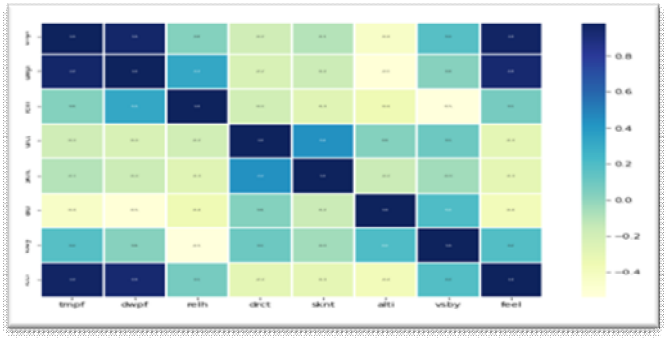


Fig.A.2. Visualizing the metar prediction for the next 24 hours.

4. CONCLUSION

In the forecasting of weather, the facebook prophet model is used for predicting the result and it gives quick and easy access.

The time series algorithms to analyse the pattern in METAR data and thereby forecasting the upcoming weather would significantly improve the efficiency of the prediction.

5. FUTURE ENHANCEMENT

The traditional model like ARIMA and SARIMA both of the algorithms takes the previous data and predicts the future.in case of SARIMA similarly takes past data. But comparing with other, the FACEBOOK PROPHET model is **explicable.**

A person without previous experience or in-depth knowledge in time series modelling can work around. Facebook's prophet is accurate and it is the fastest algorithm to predict the data.so, the facebook prophet is used for the metar forecasting.

Accurate weather forecasting helps pilots to make decision whether to land and helps them plan the most efficient and safe route,

It also helps in minimizing the impact of weather on passenger comfort and helps ensure that passengers arrive at their destination on time and in good condition.

6. REFERENCES

[1] Time series analysis with Facebook Prophet using the covid-19 data by Mitchell Krieger at Feb 20, 2021. https://towardsdatascience.com/time-series-analysis-with-facebook-prophet-how-it-works-and-how-to-use-it-f15ecf2c0e3a

[2] Emir Zunic, Kemal Korjenic, Kerim Hodzic and Dzenana Donko. Prediction on sales forecasting based on Real-World data. International Journal of Computer Science & Information Technology (IJCSIT) Vol 12, No 2, April 2020.

[3] Jan G De Gooijer, Rob J Hyndman. 25 years of time series forecasting. Journal of forecasting 1982-1985; International Journal of Forecasting 1985-2005.

[4] Kulendran, N., & King, M.L. (1997). Forecasting international quarterly tourist flows using error-correction and time-series models. International Journal of Forecasting, 13, 319–327.

[5] Priya Narayanan, Ashoke Basistha, Sumana Sarkar, Kamna Sachdeva, Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India, C. R. Geoscience. 345 (2013) 22–27.

[6] Meteorologist uses some tools to forecast the weather refer the link. https://www.noaa.gov/stories/6-tools-our-meteorologists-use-to-forecast-#:~:text=Supercomputers&text=Observational%20data%20collected%20by%20doppler,forecast%20guidance%20to%20our%20meteorologists

[7] Time series Facebook Prophet Model and Python for COVID-19 Outbreak Prediction by Mashael Khayyat, Kaouther Laabidi, Nada Almalki and Maysoon AI-Zahrani.Accepted :06 January 2021.

[8] Forecasting at Scale by Sean J.Taylor and Benjamin Letham.

[9] Thin Thin Swe1, Phyu Phyu1, Sandar Pa Pa Thein. Weather prediction model using random forest algorithm and apache spark. International Journal of Trend in Scientific Research and Development (IJTSRD) Volume 3 Issue 6, October 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470.

[10] Lorenzo Menculini, Andrea Marini, Massimiliano Proietti, Alberto Garnei, Alessio Bozza, Cecilia Moetti, Marcello Marconi. Comparing prophet and deep learning to ARIMA in forecasting wholesale food prices.

❑❑❑

VSRD International JOURNALS

A Research Division of Visual Soft India Pvt. Ltd.

REGISTERED OFFICE

154, Tezabmill Campus, Anwarganj, KANPUR - 208 003 (UP) (INDIA)
Mb.: +91 98999 36803 || Web.: www.vsrdjournals.com || Email: vsrdjournal@gmail.com

MARKETING OFFICE

340, First Floor, Adarsh Nagar, Oshiwara, Andheri(W), MUMBAI - 400 053 (MH) (INDIA)
Mb.: +91 99561 27040 || Web.: www.vsrdjournals.com || Email: vsrdjournal@gmail.com