



FPT Education

FPT UNIVERSITY

**CONFERENCE
PROCEEDINGS**



**FPT UNIVERSITY STUDENT RESEARCH CONFERENCE
PROCEEDINGS**

FPT UNIVERSITY

**Volume 3
December 2017**

FPT UNIVERSITY

FPT UNIVERSITY STUDENT RESEARCH CONFERENCE

PROCEEDINGS

**Fall Semester 2017
Volume 3**

**December, 2017
Ho Chi Minh, Viet Nam**

Table of Content

Introduction.....	i
Organizers, Faculty advisors, Program Committee	ii
Conference Program.....	iii
Keynote Speech 1: How to write a Research paper for International Publishing	
Huynh Cong Viet Ngu	1
Keynote Speech 2: Pollia: A New Approach of Web Application for Tourism Service	
Nguyen Anh Luan, Tran Phuoc Tien, Mai Duc Thang	4
Papers of Technical Sector	13
A Framework for Clustering Time Series Data	
Vuong Chanh Quyen, Trinh Thanh Quyen, Nguyen Minh Nhut.....	15
Applying iBeacon Technology to build Automation Tollbooth Station: Combined with vehicle number plate recognition technique	
Le Vuong Quoc Huy, Nguyen Chi Hieu, Pham Bao Tin, Do Viet Son	21
Apply Collaborative Filtering to give Suggestions for Users in Web Application	
Nguyen Viet Cong Hau, Luu Nguyen Ngoc Thu, Nguyen Ngo Duc Khai.....	27
Building Crowd Mornitoring of Expo System using iBeacon Technology	
Ngo Quang Tuan Anh, Vo Thanh Danh, Nguyen Ngo Duc Khai	31
Customer Relationship Management System for QTSC	
Nguyen Thanh Tung, Huynh Huu Nhan, Nguyen Thi Phuong Nhi, Tran Nguyen Bao Duy	33
DDoS Attack Prevention System	
Ma Lam Ngoc, To Thanh Thuyen, Nguyen Hoang Tuan	36
Dendrobium Orchid Care Solution: Delaying in small and medium system to support people who like taking care of orchids at their home without high cost	
Nguyen Quang Huy, Nguyen Tuan Sang, Mai Thanh Hieu, Huynh Quang Dat, Tran Hong Son.....	40
Develop a Security Information Event Management System	
Le Xuan Minh, Vo Thong Minh, Tran Quoc Phuong Nam, Do Duy Thinh	46
Door Access and Payroll Monitoring System: Delaying in small and medium company require a system to manage staff access	
Nguyen Xuan Huy, Bui Cong Khanh, Bui Quang Vien, Nguyen Xuan Huy	52
Electronic Devices B2B Website	
Nguyen Hoang Phuc, Nguyen Thi Thai Han, Nguyen Van Bach, Phan Nhat Minh.....	57
Football Field Reservation System	
Mai Minh Quy, Truong Huu Thanh, Phan Minh Huan, Pham Trung Hieu	61
Hacking and Security Techniques on Web Applications	
Ho Bao Tien, Hoang Quoc Hung, Lam Ngo Quyen, Huynh Thanh Trung	65
Intent Analyzer, Dialog flow Controller and Sentence Suggestion: Location-based Chat bot team	
Tuong Duc Khang, Nguyen Van Hung, Duong Hai Quang, Thai Hoang Duy Bao	70

sProcurement: TOPSIS Method for Supplier Selection Problem

Hoang Vinh Son, Nguyen Dinh Thong, Vu Thai Nam, Huynh Ngoc Dat 77

University Admission Counseling System for High School Students

Ngo Nguyen Thuy Van, Le Thanh Danh, Vo Manh Hung, Nguyen Le Minh 81

**Warning System on Parameters of Devices in the Family: Deploying warning system for the family
in Vietnam**

Vo Phuc Hai, Ton That Minh Tri, Nguyen Phu Ngoc Trai, Nguyen Kim Cuong 84

Papers of Economic Sector..... 95

Analyzing the Consumer Satisfaction for an Internet Service Provider

Vo Thi Tuyet Nhi, Le Thi Diem Xuan, Vo Quang Sang, Pham Anh Duc 97

Analyzing the Customer Satisfaction in Using Websites that Market Digital Products and Services

Dao Manh Cuong, Nguyen Pham Bao Khanh, Tran Cong Hau, Do Hoang Tuan 103

Analyzing the Influence Factors to Consumer Acceptance of Internet of Things Technology

Pham Quoc Trung, Nguyen Thanh Hoai Ngan, Ho Thi Dieu Huyen, Phung Mai Khanh 106

Conference Participants 109

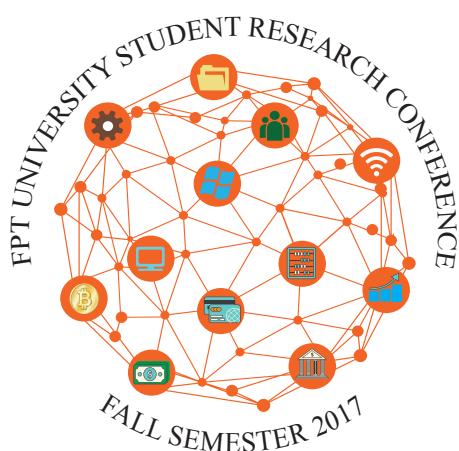
Awards & Photos Appendix 113

About the Proceedings

The Student Research Conference Proceedings is periodical publication follow every semester of FPT University.

The Proceedings gather all the papers of the lecturers, researchers in the conference, and the main content is the student's subject papers had defended in front of the scientific council, with a scientific content highly actual, acceptable post in the Proceedings.

At present, The Student Research Conference Proceedings was published by English language. This is published for internal circulation only. Readers could be found it in the Library Department of HCMC FPT University and Ha Noi FPT University, also found it with electronic version on Dspace at: <http://ds.libol.fpt.edu.vn>



Introduction

Students' scientific research is intellectual activities very important that support students to apply methodologies and methods of scientific research in studying and practicing; since, using a combination of learned knowledge to conduct research-based cognitive activity, contributing to solving scientific problems in real life and occupation. Recognizing scientific research is one of important tasks that FPT University has been step by step building, deploying and encouraging scientific research for staff, lecturers and students since the first day of establishment.

As a result of fact, The following success of The Student Research Conference before, The Student Research Conference - Fall Semester 2017 (3rd) which was held in December 2017, which the purpose of bringing environments for students to report their results of research to the Scientific Research Council; contributing to encouraging and promoting scientific research activities as well as applying results of research in reality as soon as possible. The conference is the place to receive the comments and suggestions from Scientific Research Council, lecturers and staffs to improve movement and quality in students' scientific research. Throughout this conference, FPT University hopes to promote the students' scientific research which also to raise awareness of the importance of science in improving the quality of teaching and training. In this conference, there are 20 groups with 20 different subjects belonging to three faculties: Software Engineering, Information Assurance and Business Administration. The Scientific Research Council has based on an evaluation process in order to select 7 topics, with strict criteria to be presented in The Student Research Conference - Fall Semester 2017 (3rd).

The organizers sincerely appreciate the interest of Executive Board, Scientific Research Council, Head of Departments, Lecturers, Staffs and all students whom have contributed for the success of the Conference. We are thankful to Mentors who have guided students their scientific research. We would like to receive all comments and suggestions in order to develop and improve the quality of scientific research in the future.

Organizers

Tran Ngoc Tuan, Ph.D. (Vice Rector of FPT University)
Than Van Su, M.E (Head of Academic Affairs Department, HCMC FPT University)
Kieu Trong Khanh, M.E (Head of ITS Department, HCMC FPT University)
Vo Minh Hieu, MBA. (Head of Business Faculty, HCMC FPT University)
Ngo Dang Ha An, M.E (Head of IA Department, HCMC FPT University)
Dinh Truong Lam (Head of Library Department, HCMC FPT University)
Pham Trong Phuong Tra (Library Officer of Library Department, HCMC FPT University)
Nguyen Manh Hung (Library Officer of Library Department, HCMC FPT University)

Faculty advisors

Than Van Su, M.E (Head of Academic Affairs Department, HCMC FPT University)
Kieu Trong Khanh M.E (Head of ITS Department, HCMC FPT University)
Nguyen Duc Loi, M.E (Lecturer of IT Department, HCMC FPT University)
Nguyen Huy Hung, M.E (Head of SE Department, HCMC FPT University)
Ngo Dang Ha An, M.E (Head of IA Department, HCMC FPT University)
Lai Duc Hung, M.E (Lecturer of IT Department, HCMC FPT University)
Nguyen Sieu Dang, M.E (Lecturer of IT Department, HCMC FPT University)
Lam Huu Khanh Phuong, M.E (Lecturer of IT Department, HCMC FPT University)
Vo Minh Hieu, MBA. (Head of Business Faculty, HCMC FPT University)

Program Committee

Nguyen An Te, Ph.D. (Lecturer of Technical Sector, FPT University)
Than Van Su, M.E (Lecturer of Technical Sector, HCMC FPT University)
Nguyen Trong Tai, M.E (Lecturer of Technical Sector, HCMC FPT University)
Nguyen Quoc Cuong, MBA. (Lecturer of Economic Sector, HCMC FPT University)
Nguyen Ngoc Danh, MBA. (Lecturer of Economic Sector, HCMC FPT University)
Nguyen Quang Dung, MBA. (Lecturer of Economic Sector, HCMC FPT University)

Editor

Nguyen Huy Hung, M.E (Head of SE Department, HCMC FPT University)

Conference Program

Saturday, December 23rd, 2017

- 07:00 - 07:30** Welcome Delegates and Guests
- 07:30 - 07:40** Introduction of conference opening - Speech of Tran Ngoc Tuan, Ph.D., Vice Rector of FPT University
- 07:40 - 08:00** Keynote Speech 1: How to write a Research paper for International Publishing
Huynh Cong Viet Ngu, M.E (Lecturer of IT Department, HCMC FPT University)
- 08:00 - 08:20** Keynote Speech 2: Pollia: A New Approach of Web Application for Tourism Service
Presented by Nguyen Anh Luan, Tran Phuoc Tien, Mai Duc Thang

Reports of Students from Economic Sector

- 08:20 - 08:40** Analyzing the Consumer Satisfaction for an Internet Service Provider
Presented by: Vo Thi Tuyet Nhi, Le Thi Diem Xuan, Vo Quang Sang, Pham Anh Duc
- 08:40 - 09:00** Analyzing the Influence Factors to Consumer Acceptance of Internet of Things Technology
Presented by: Pham Quoc Trung, Nguyen Thanh Hoai Ngan, Ho Thi Dieu Huyen, Phung Mai Khanh

09:00 - 09:15

Tea-break

Reports of Students from Technical Sector

- 09:15 - 09:35** Hacking and Security Techniques on Web Applications
Presented by: Ho Bao Tien, Hoang Quoc Hung, Lam Ngo Quyen, Huynh Thanh Trung
- 09:35 - 09:55** Applying iBeacon Technology to build Automation Tollbooth Station
Presented by: Le Vuong Quoc Huy, Nguyen Chi Hieu, Pham Bao Tin, Do Viet Son
- 09:55 - 10:15** Football Field Reservation System
Presented by: Mai Minh Quy, Truong Huu Thanh, Phan Minh Huan, Pham Trung Hieu
- 10:15 - 10:35** Warning System on Parameters of Devices in the Family
Presented by: Vo Phuc Hai, Ton That Minh Tri, Nguyen Phu Ngoc Trai, Nguyen Kim Cuong
- 10:35 - 10:55** University Admission Counseling System for High School Students
Presented by: Ngo Nguyen Thuy Van, Le Thanh Danh, Vo Manh Hung, Nguyen Le Minh
- 10:55 - 11:05** Statements for closing Conference, Speech of Mr. Than Van Su (M.E) - Head of Academic Affairs Department, HCMC FPT University
- 11:05 - 11:15** Awards Ceremony

Keynote Speech 1:

How to write a Research paper for International Publishing

Huynh Cong Viet Ngu, M.E

Lecturer of IT Department, HCMC FPT University

Abstract

Nowadays, in addition to focusing on developing and improving the quality of training, most universities in Vietnam are increasingly interested in another important activity for student -Scientific research. One of the most primary tasks of the scientific researcher in general, and the students, who are participating in this kind of activity is the communication of technical results to the broader scientific community. Whether in written or oral form, scientific communication is a critical step in the scientific method and is the key driver of movement within a scientific field. Therefore, the construction of a written scientific manuscript must not be taken lightly. For this reason, in this article, I propose one approach for writing a research paper for international publishing specially.

Keywords

*Student research; International publishing;
Research paper construction*

I. INTRODUCTION

Currently, with the policy of encouraging students to participate in scientific research activity at most universities, it can be said that students receive a lot of benefits from this activity. Typical benefits can be grouped into two main groups. First, a significant increase in knowledge and research experience. This activity requires the researcher have to continually update their knowledge, so finding and reading more material are essential. Through this, the research skills as well as knowledge for their topic will increase. In addition, students have the opportunity to work with instructors so they will be more thoroughly instructed on their research issues.

Secondly, scientific research helps students enhance the skills needed for their future work and life such as critical thinking skill, time management skill, teamwork, computer techniques, in which the most important is critical thinking skill so they can evaluate the problem in many aspects. In that context, the publication of research results or, in other words, the construction of a written scientific manuscript according to international standard is

one of the primary tasks of the scientific researchers in general, and the students in particular so it should to be considered in the right way. For this reason, in this article, the structure as well as a approach for writing a research paper for international publishing are proposed to help students have a better view on science articles.

The rest of the paper is organized as follows: Section II gives a structure as well as a approach for writing a research paper for international publishing and the last section, section III, is my conclusion

II. RESEARCH PAPER: STRUCTURE AND APPROACHES

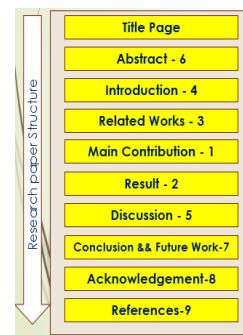


Figure1. Research paper structure

A primary task of a researcher is to report the technical results to the scientific community to make it broader. Whether in written or oral form, scientific communication is a critical step in the scientific method and is the key driver of movement within a scientific field. Therefore, the construction of a written scientific manuscript must not be taken lightly.

Manuscripts that are successfully submitted to a journal for publication have three main components: overall idea, research process, and the presentation result. It should be noted that a poor idea or a poorly designed investigation can not be saved by an excellent presentation of the work, and equally an excellent idea that is well investigated can still be doomed by a poor presentation. For this reason, a research paper structure and a reasonable approach to writing a scientific manuscript are shown in Figure1.

First write the Main Contribution section, that is

your research, and then write the Results section. Reconsider the scientific questions the manuscript will address, again referring to your research protocol, and then write the Introduction and Related Works. Next, combine the Introduction, Related Works and Results to write the Discussion. Summarize everything in an Abstract, and then refocus the Abstract into a Conclusions section. Below is a brief discussion of each of the sections. These are only suggestions on how a scientific manuscript may be written. In general, the purpose of a scientific manuscript is to construct a clearly written document that describes a question and then logically presents an answer to this question that is based upon theoretical or experimental results.

1. Title Page

A title page should be included the title of the manuscript, which should be short and simple, as well as authors. Provide approximately 3-6 key words. Finally, provide complete contact information for the corresponding author. Authorship and the order of authorship must be agreed upon by all of the authors. Example is shown in Figure2

B+-tree Construction on Massive Data with Hadoop
Huynh Cong Viet Ngu*
* Department of IT, Lecturer of FPT University, Socialist Republic of Vietnam
nguhcv@fe.edu.vn

Figure2. Title Page Example

2. Abstract

The abstract is typically a single paragraph. The abstract should be considered as an independent document, so that the abstract does not rely upon any material in the body. The first sentence should clearly state the objective of the research. If the research is based upon a hypothesis, it should be stated and followed with statements describing its basis and evaluation. The subsequent sentences describe how the investigation was carried out. The following sentences describe, with as much precision as possible the results of the research. The final sentences describe the significance of the results and the impact of this work on the field of study. Example is shown in Figure3

B+-tree Construction on Massive Data with Hadoop

Huynh Cong Viet Ngu*

* Department of IT, Lecturer of FPT University, Socialist Republic of Vietnam
nguhcv@fe.edu.vn

Abstract. The data processing in the Socialist Republic of Vietnam (Vietnam, hereunder) is in an early stage and a variety of problems are needed to be solved. In the Vietnamese banking and financial sectors, where managing and storing of customer data and transaction histories are being emphasized as never before, the volume of data to be secured on a daily basis are explosively increasing due to rapid economic development so that the relevant authorities are seeking an efficient and reliable way to manage them. Being a widely known popular variation of B-tree, B+-tree is considered as a most adequate tree-type data structure for bulk data. Nevertheless, as it is quite time-consuming to construct a B+-tree for massive data the authors propose a Hadoop framework-based parallel B+-tree system to deal with the problem. The system is largely divided into 3 phases. First, data are partitioned and distributed evenly such that each partition will have almost the same amount of data volume. Second, a parallel local B+-tree system is constructed. Finally, some small-scale B+-trees are constructed and integrated into the complete form of B+-tree which will be dealing with an entire data set. The authors expect that the proposed system will offer an efficient index structuring while reducing data processing time.

Keywords: B-tree, B+-tree, Hadoop, Map-Reduce, Big Data, Cloud Computing.

Figure3. Tille Page, Abstract and Keywords Example

3. Introduction

The introduction is a short review of the literature pertain to the research topic, starting with broad topics and slowly focusing on the work at hand. One approach may be to start with one or two paragraphs that introduce the reader to the general field of study. The subsequent paragraphs then describe how an aspect of this field could be improved. The final paragraph is critical. It clearly states, most likely in the first sentence of the paragraph, what experimental question will be answered by the present study. The hypothesis is then stated. Next, briefly describe the approach that was taken to test the hypothesis. Finally, a summary sentence may be added stating how the answer of your question will contribute to the overall field of study.

4. Related Works

The section that often follows the introduction is the related work section. The related work section may also be called a literature review. The point of the section is to highlight work done by others that somehow ties in with your own work. It may be work that you're basing your work off of, or work that shows others attempts to solve the same problem. There are a couple of schools of thought when it comes to writing the related work section. The first school of thought is to mention absolutely everyone and every paper that may be even remotely related to the topic at hand. Personally, I don't like this method. The second is to pick and choose the most important papers and to only talk about them. If you are going to mention a paper, it needs to have some concrete relation to the work you're doing. You should be able to (easily) defend each paper as to why you chose it and why it's important. Some papers will be easy to defend – these are the papers that present work you're building upon or papers that are about other

solutions to the problem you're working on. Others are more difficult. If it's difficult to defend, then ask yourself why you're including the paper and if it's really necessary.

There are some arguments that, the related works should be a part of the introduction section. However, in my opinion, it should be a separate part. We have two benefits in doing so. First, the reviewer can better understand the work we are doing, the second, helps the reviewers appreciate our research capabilities.

5. Main Contribution

This section should be a straightforward description of the methods used in your study. Each method should be described in a separate section. Begin, in a single section, with a statement of the materials used in the study, indicating the vendor and vendor contact information for each material. This information is critical so that readers have the capability to repeat the work in their own institutions.

Next describe, in separate sections, each key procedure and technique used in the study. Keep explanations brief and concise. Similarly, if a theoretical or modeling component is utilized, it should also be incorporated in the initial portion of the Methods. Finally, remember to describe the statistical analysis methods that were utilized to analyze the results

6. Result

The Results section presents the experimental data to the reader, and is not a place for discussion or interpretation of the data. The data itself should be presented in tables and figures. Introduce each group of tables and figures in a separate paragraph where the overall trends and data points of particular interest are noted. You may want to indicate the placement of a particular table or figure in the text. Note that each table and figure in the paper must be referred to in the Results section. Be succinct

7. Discussion

The discussion section, often the most difficult to write. Begin this section with a brief paragraph that again gives an overview to the work. Summarize the most important findings and, if applicable, accept or reject the proposed hypothesis. Next, identify the most interesting, significant, remarkable findings that were presented in the Results section, and contrast these findings in light of other studies reported in the literature. Finally, at the end of the Discussion section, consider the other works in the literature that address this topic and how this work contributes to the field of study

8. Conclusion

Again, first introduce the work and then briefly state the major results. Then state the major points of the discussion. Finally, end with a statement of how this work contributes to the overall field of study.

9. Acknowledgement

Provide a brief statement acknowledging the efforts of any participants or consultants who are not included as authors of the manuscript. List all of the funding sources for the research

10. References

Include all references that have been cited in the text. The references should be well considered, so that they contain all key sources in the field as well as previous studies that support or motivate the present work. However, do not include extraneous references in an effort to simply cite particular authors or journals. You must use the reference format that is mandated by the journal to which you are submitting the manuscript. There are some software packages make citing easy (Mendeley,...)..

III. CONCLUSION

A scientific manuscript is meant to convey technical information to the reader. Therefore, it is generally designed to be a straightforward presentation and discussion. Paragraphs and sentences should be simply constructed. One point of view that supports this concept is that the scientific aspect of the manuscript may be challenging enough for the reader to comprehend, therefore the text itself should support to convey the scientific information, rather than acting to further obscure the concepts and results.

REFERENCES

- [1] <https://writing.wisc.edu/Handbook/PlanResearchPaper.html>
- [2] <https://www.wikihow.com/Write-a-Research-Paper>
- [3] Huynh Cong Viet Ngu, "B+tree construction on Massive data with Hadoop" (Cluster Computing – DOI 10.1007_s10586-017-1183-y).

Keynote Speech 2: Pollia: A New Approach of Web Application for Tourism Service

Nguyen Anh Luan
FPT University
HCMC, Viet Nam

Tran Phuoc Tien
FPT University
HCMC, Viet Nam

Mai Duc Thang
FPT University
HCMC, Viet Nam

Abstract

Nowadays, people are so familiar with traveling by tour are provided by travel companies, which leads to limited places and time constraints. Therefore, the trend of backpacking tourism is emerging to attract people especially young people. Backpacking tourism is a kind of traveling which people will choose places to visit, time to spend, how to get there, eating as well as stay. Since everything must be planned on the location of the destination, the traveler must prepare the information, the price as well as the experience of the person who went. Most of the applications in Vietnam are very diverse and serve many comparative and search features, but those applications still do not meet the needs mentioned above. For that reason, we have created the Pollia application to provide information about places, events, places and events, and other user-contributed suggestions from which users can choose. Pick and connect the places and events they like to make their own trip, and they can save that trip for the experience. In addition, the efficiency of querying information from the user is also our concern in the right way.

Keywords

Pollia; Tourism service; Tripbook; Self-made tour.

I. INTRODUCTION

In the context that Vietnam has been integrating with the world through signed economic and trade agreements as well as Decision No. 2522/QDBVHTTDL issued and effective on July 13 2016: “to focus on promoting cultural tourism products in association with studying life, understanding food, craft villages, festivals and cultural heritages.

Focusing on the experience, diversity and richness of cultural values to form each product are highly attractive ...”; The economic life of the people in Vietnam has been improved, and the spirit of the state is more concerned about tourism. This shows the need for a solution to the problem of tourism in the technology era.

Tours provided by travel companies are often scheduled, which will make visitors less likely to experience the trip. It is limited by locations and time. There are many good applications for searching information, prices and experience sharing of locations such as: CococMap, Foody, Here We Go, Trivago, vnTrip. These applications make it easy for users to find information and comment on places. However, they do not provide a solution to connect places to create self-made trips.

Pollia was first introduced in June 2017 with the purpose of introducing and connecting tourism destinations. Information about places we collect from online sources, is analyzed and stored in the database to serve the search needs of users. Data is updated periodically on information collected by the system or by users contributing to enrich the data source of information, forms of tourism, culture of places, events in Vietnam. When users visit Pollia’s website, they can search for information about places and events they are interested in. The system will process the search, return the result and display the results in the selected user interface. Users can add points and events into their Tripbook. Tripbook is a collection of places and events added by the user and a user can create many different Tripbooks. It is important here that users can make their own trips to destinations with clear information and manually schedule their trips. In terms of data, in addition to SQL, Pollia applies GraphDB to increase DB performance to increase user experience.

II. RELATED WORK

1. Current Applications

1.1. Foody

In 2011 through mid-2012, the Vietnamese market exploded in group buying pages. However, due to the failure of food information sites in Vietnam,

there is a need to exploit this type of website. Launched in 2012, Foody brings an online restaurant search and rating services in Vietnam. Currently, Foody has 10 million monthly visitors based on 5 million users and more than 100 thousand of reviews.

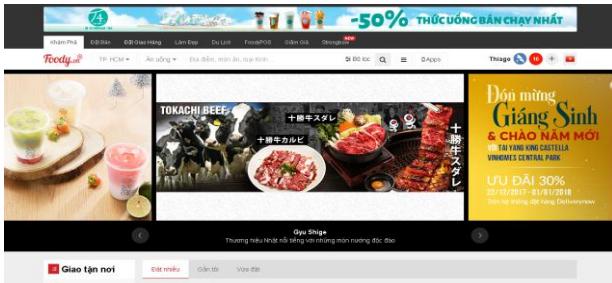


Figure 2.1.1: Foody.vn

Foody does not operate under the forum model and reviews are long, which brings a culinary social network to the user. Often the difficulty faced by the culinary websites is that the user does not have a habit of sharing and evaluating, resulting in a lack of objective evaluation information and limited information about restaurants. Not letting Foody follow the path is the most up-to-date information from forum models, a culinary social network that provides us with an unlimited source of information from our users. With the users receiving reward points from reviews and providing information to convert vouchers, Foody has encouraged users to become familiar with Foody. When there is data that the user evaluates and supplied, Foody applies the optimization to the SEO. Food for the trustworthiness of yourself and the cafeteria in the basis of the user evaluation is objective from the very Much appreciated by different users. With Foody becoming popular, eateries are well known and many other eateries will want to post their information at Foody. They can promote, receive feedback from customers and get more out of these responses. Foody has also provided two new services, tablenow and deliverynow, based on food data. They are the first page bookmakers to offer more offers to the eateries associated with themselves.

Foody has received an unlimited data source from users and restaurants. Their success is to know how to put themselves as the intermediary link supply-demand of the restaurant and diners. They have created a win-win relationship, bringing benefits to them and diners. Foody has the information to supplement and update from users and eateries, eateries that are promoted and feedback as well as their coverage from foody and users, and users gain information. Trust the food they need as well as rewards from the review.

However, if they want to travel to another area and

look for restaurants, hotel and another kind, they need to put them in a plan for their travels. A digital map will further assist them in finding places. And this is the point we are pushing to bring a travel service based on search services like foody.

1.2. Coc Coc Maps

With the success and benefits of the food information site, CocCoc has decided to expand not only food but many other services such as hotels, and spas. POIs (Point of Interest) brings a project named CocCoc Maps. CocCoc Maps provides a digital map service for the Vietnamese market with the goal of providing users with the most convenient and up-to-date digital map.

Starting in 2013, CocCoc Maps collects 500,000 service locations. Difficulties emerged when the data collection was significantly changed by the rate of change of enterprises in Vietnam, especially small and medium enterprises. According to statistics, 40% of POIs have changed from 2013-2015.

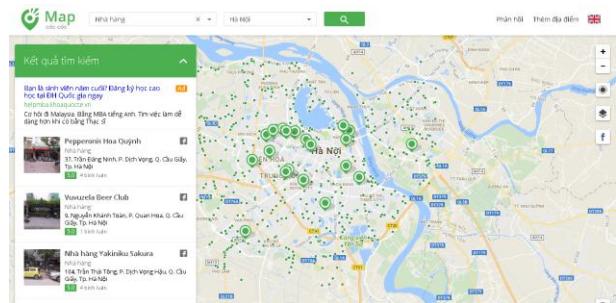


Figure 2.1.2: Coc Coc Maps

CocCoc changed its way of capturing information from manual capture to automatic recording using a dedicated camera on a motorcycle with a frequency of 2.3 months in Hanoi and Ho Chi Minh City, one month in Da Nang ... Along with that they develop synchronous tools, extract images corresponding each location collected from the camera. Efficiency from this change brought great. Updates from 50 POIs / h to 500 POIs / h. And the collection and updating sites are up to 114,000 POIs per day. These changes have been processed, updated efficiently, speedily and at the same time CocCoc Maps is also crawling from other sources on the Internet to avoid missing, duplicate or confusing places. CocCoc Maps is rolling out more user ratings and reviews of POIs aimed at updating information. The success of CocCoc Maps comes from updating information and the accuracy of locations. They have surpassed Google Maps for the accuracy of service locations in Vietnam by collecting their information. These are the highlights of their position on the digital map service. in Vietnam market.

CocCoc Maps is so powerful, but if not just looking for places that make up a plan, a tour based on linking

places they need to find on digital maps, It will be more relevant to the needs of the people, which will be more beneficial from the amount of data and maps they bring. Furthermore, this is something that we will add and rely on. this powerful digital map idea.

2. From RDBMS to NoSQL

Relational Database has been a successful data storage technology for the last 20 years, ensuring data sustainability, concurrent data processing, and unity of storage and retrieval. However, with the explosion of digital data today, with websites with millions of users, data binding and archiving becomes more difficult. Many data needs make many types of data need to store different types of information, are not consistent with each other and do not follow a common structure.

NoSQL becomes the choice to solve this problem and provides different ways of managing data to store them. NoSQL is meant to be a more advanced version of RDBMS, which can store not only simple types of data types, but also data types such as arrays, video, images, etc. The importance of NoSQL Again, Schemaless, there is no common structure for data types; the need to add attributes to data is to add them without regard to the structure of the remaining records. In addition, the data retrieval of NoSQL is easier with the organization that it provides.

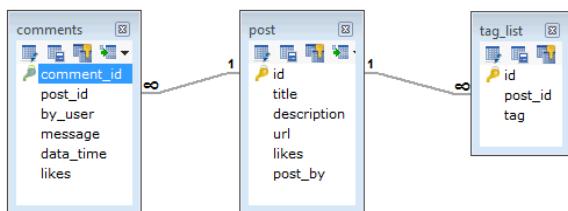


Figure 2.1: Organizes data of 2 tables Post and Comment in 1 Forum by RDBMS

Looking at Figure 2.1, we will see if we want to display a full post, we have to use the JOIN function 3 times and browse 3 tables. This results in reduced access speed.

With NoSQL like Figure 2.2, we only need one record to store all the required data for a full post. With this storage, the amount of data to browse will be less and the speed of access will be faster.

NoSQL comes in a variety of formats to support different types of storage and different data representations supporting access. The most common of these is the Document Database, which is used to store data in the form of BSON / JSON / XML and Graph Database, used to store and link data as a graph. We will talk more about the advantages of Graph Database, the type of NoSQL we apply to our application.

```

{
  "_id: POST_ID
  title: TITLE_OF_POST,
  description: POST_DESCRIPTION,
  by: POST_BY,
  url: URL_OF_POST,
  tags: [TAG1, TAG2, TAG3],
  likes: TOTAL_LIKES,
  comments: [
    {
      user:'COMMENT_BY',
      message: TEXT,
      dateCreated: DATE_TIME,
      like: LIKES
    },
    {
      user:'COMMENT_BY',
      message: TEXT,
      dateCreated: DATE_TIME,
      like: LIKES
    }
  ]
}
  
```

Figure 2.2: Organization of 2 tables Post and Comment data in a Forum using NoSQL (Document DB)

Figure 2.3 shows the example of organizing the data in the form of a Graph Database. Because the current data is linked to each other, an object can have multiple links to and go to different objects. Relationships are so overwhelming that we can not express a single record as a Document Database. And that's why we use Graph Database.

Graph Database uses Node and Relationship to organize its data. Each Node is any data type, which contains the properties needed for that Node itself, not a single data type. And Relationship is a way for nodes to link together expressing unlimited relationships. When you need to create a new attribute, a new type of data, a link that has never before in the database, just add it. The benefit of Graph Database from this is huge.



Figure 2.3: Graph Database

The most powerful feature of the development of NoSQL is Polyglot Persistence. It is the use of different types of databases to store different types of data to serve the nature and speed of processing of that type of data.

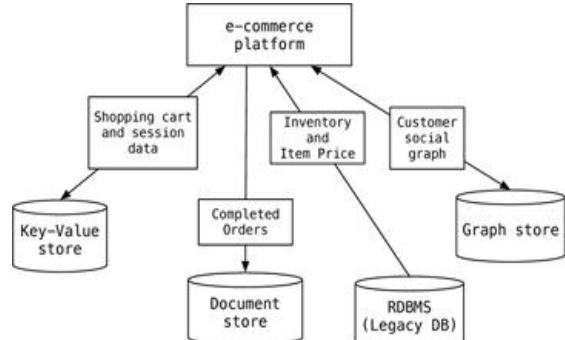


Figure 2.4: Implement Polyglot Persistence for e-commerce

Let's take an example of how to organize data for an e-commerce application. This data organization is a template for optimizing storage and retrieval speed. Shopping cart and Session Data are data that are differentiated by unique key and their data has a common structure so we use Key - Value Store

for storage. For orders, items inside the order may be different, so we use the Document Database to store it as unstructured data. Inventory and Item Price if fixed, do not change the structure we can use RDBMS to store them. And the connection between customers is very much, so the customer data organization is the Graph Database.

SQL is still used for storage for many different applications. But as the data becomes more diversified, the number of NoSQL and how to use Polyglot Persistence is a better choice for optimizing storage and retrieval. And that's why we chose Graph Database for our application.

III. POLLIA: A NEW APPROACH FOR TOURISM SERVICE

1. Existed way

Using the Web application platform, Pollia offers users a way to find information about places and events they are interested in through the Internet environment. When accessing Pollia, users can choose from two options: viewing the map and grid view. The figure 3.1.1 describe the Pollia system as diagram.

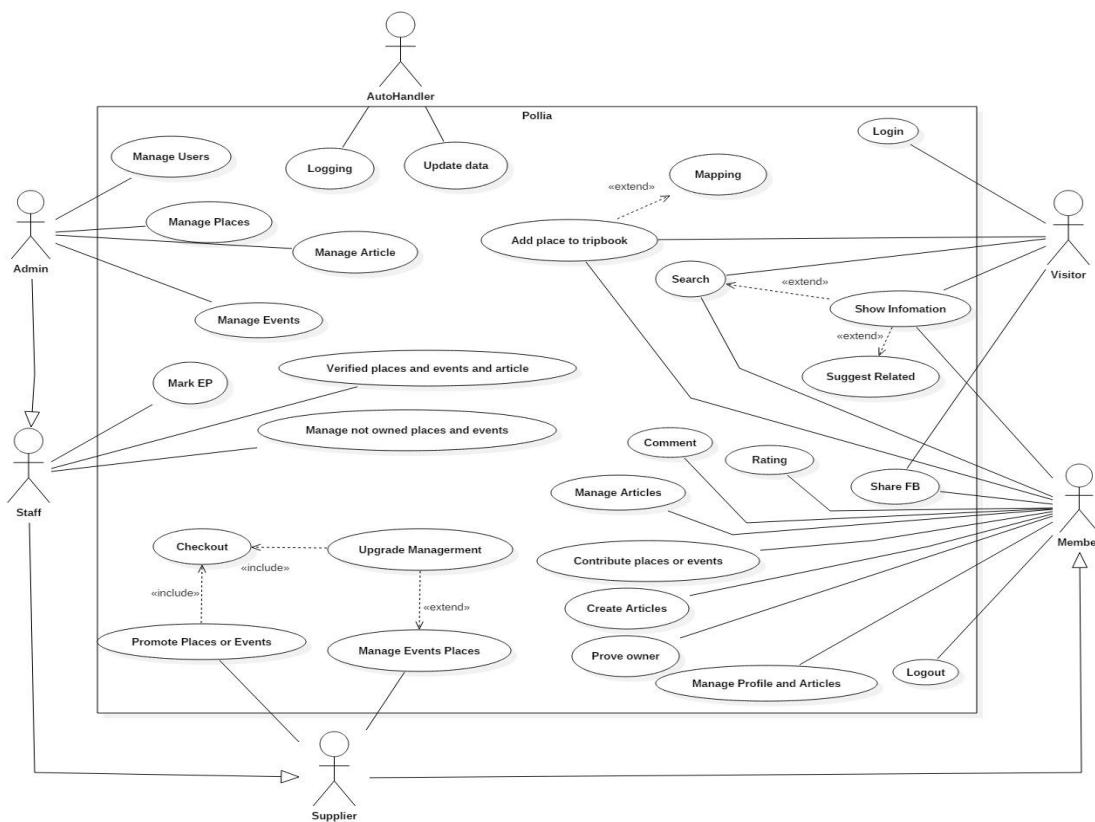


Figure 3.1.1: Pollia's usecase

- With a view of grid, users get a more intuitive view of the way and the way they move between locations. Users will enter basic information about what they want to find through search boxes and filters. The returned search results will be displayed

as a list in a frame (located on the right-hand side of the computer interface and tablet). By clicking on a location, users can see information about the place in the same frame and manipulate the location directly on the map.



Figure 3.1.2: Pollia's map view

- With grid view, users will have a more general view of the classification, area, time, and comment information about what they are looking for. Users will enter basic information about what they want to find through search boxes and filters. The results are presented in a grid pattern. By clicking on a location, users can view information about a place in a new page and interact directly with places on that page.



Figure 3.1.3: Pollia's list view

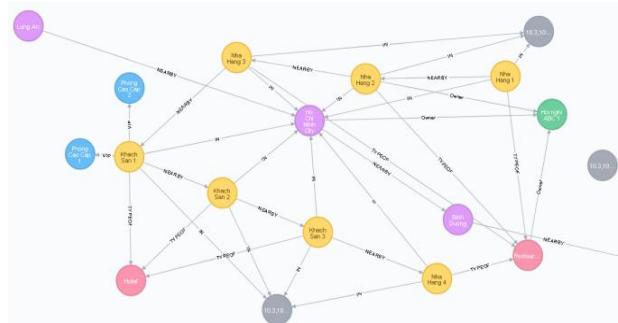


Figure 3.2.1: Implement Graph Database to Pollia Apps

Neo4j (a Graph Platform) is used for storing data and querying by Cypher Query Language. The reasons of using Graph was represented above by us. And here is how we organize data and connect them together:

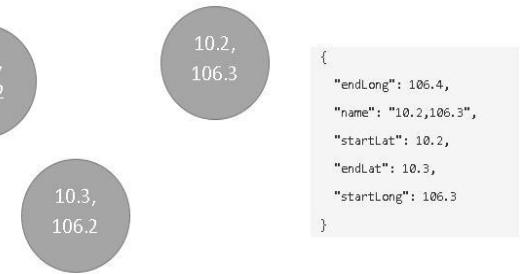


Figure 3.2.2: Scope Nodes

The important thing we aim to optimize the management of locations on a digital map is to divide the map into small squares, each with an area of $0.1 * 0.1$. By splitting the map like this, we can easily display locations in an area displayed on the user's screen by searching for corresponding scopes and finding locations that link to the scopes. There. This limits us to having to search for all the different places.

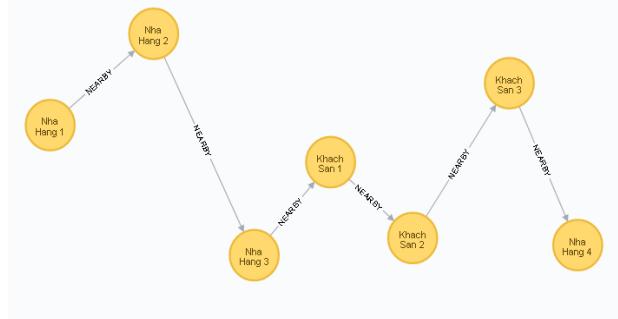


Figure 3.2.3: Place Nodes

2 Data Storage

2 Data Storage
Provides optional features for users and adds locations in their own tripbook and connects points in the tripbook to form trips. In addition, we will apply GraphDB to the system to help improve query performance.

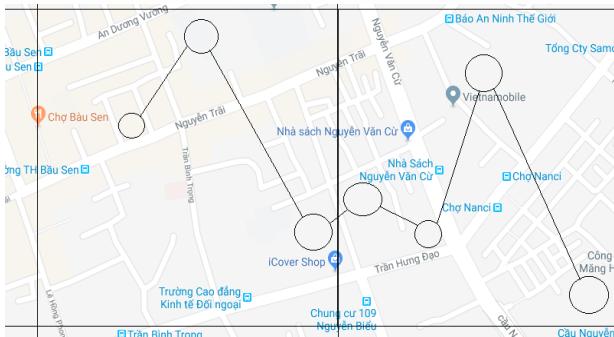


Figure 3.2.4: Place Nodes in Maps

Each Place is connected to the nearest Left Place and Place nearest to the nearest to easily find and manipulate it. The way to determine the left-hand side of a place is based on its Longitude attribute (Longitude based on geographical location). The formula for calculating the distance between two places is based on Latitude and Longitude

$$D = (A. \text{Longitude} - B. \text{Longitude})^2 + (A. \text{Latitude} - B. \text{Latitude})^2$$

D: Distance

A: Current point

B: Calculated point

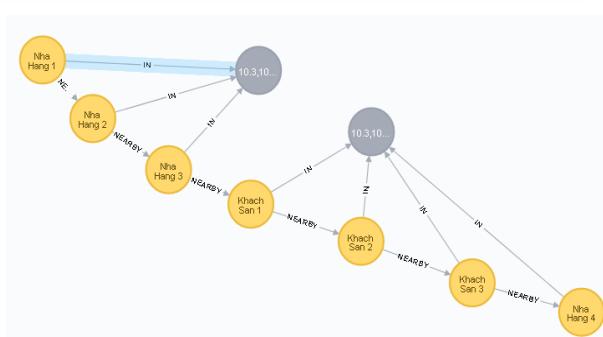


Figure 3.2.5: Place Nodes and Scope Nodes Relationship

We only link one place to the other two, right and left, to limit the relationship and match the browsing style when adding and finding places near it. When add 1 Place A, the system will find the left node L of A with the closest distance A and the link L found with A. And connect node A with node R, L connected to R. And then unlink between L and R. Figure 3.2.6 describe this one.

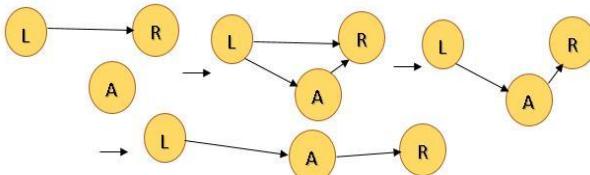


Figure 3.2.6: Add Place

This method is like dealing with variables in Linked List. We will have the same with Delete. And when we search for the closest place to the selected location, we will find the 10 nodes on the left and the 10 nodes on the right to display. When looking for types in an area, we will search under the PlaceKinds. Placements associated with PlaceKinds need to be displayed.

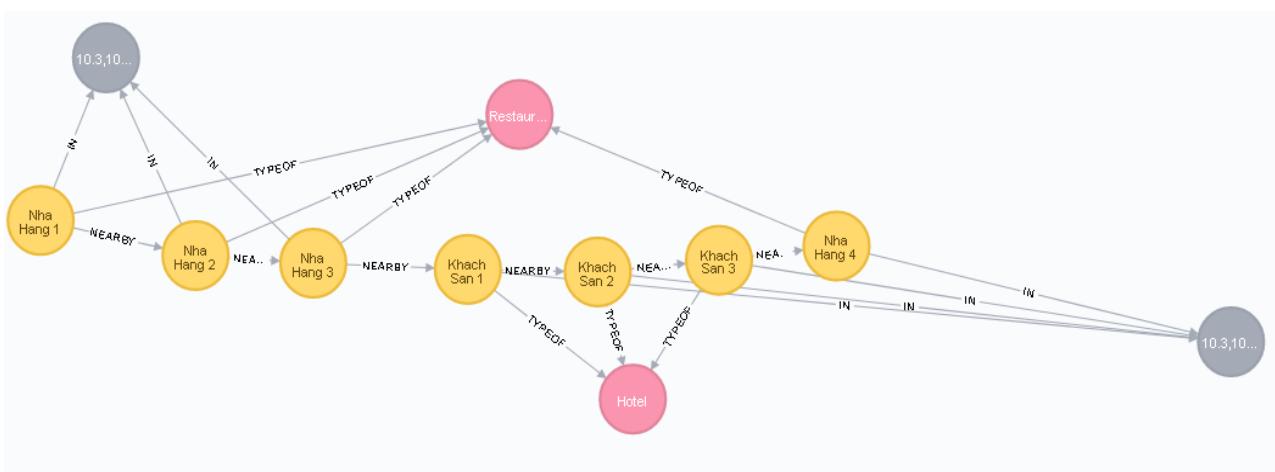


Figure 3.2.7: PlaceKind Nodes (Pink)

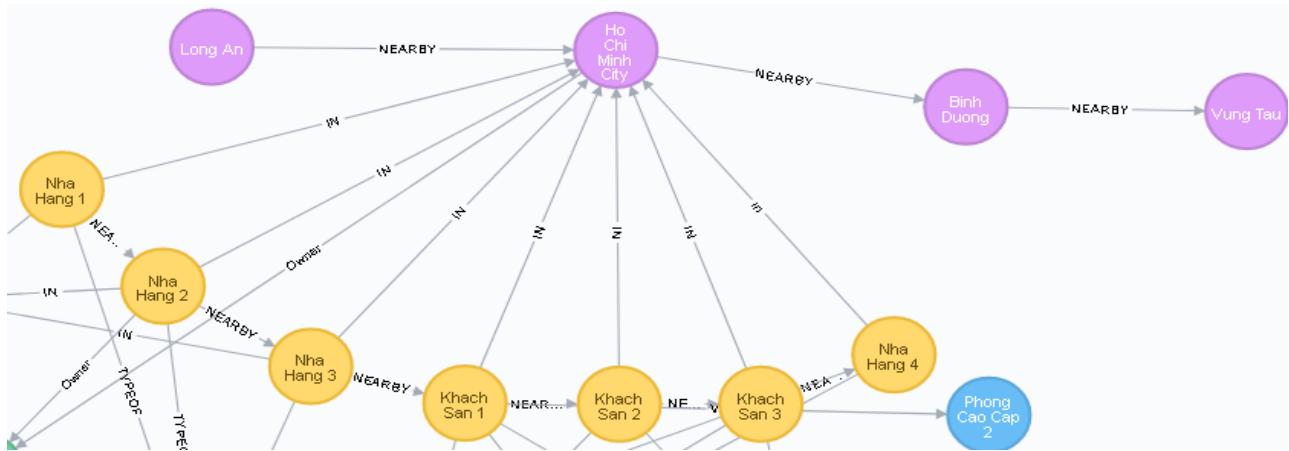


Figure 3.2.8: City Nodes

Accompanying the data is the Event at the places and genres that take place. These Place and PlaceKind will still be linked to the Event. The images are also organized similarly described in figure 3.2.9.

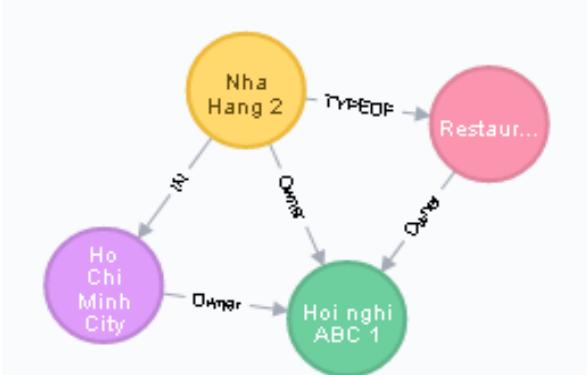


Figure 3.2.9: Event Node and Relationship

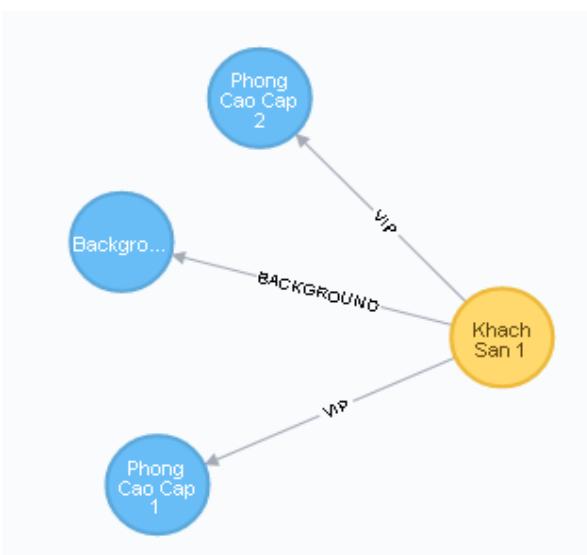


Figure 3.2.10: Image Node

When we want to add any object type, we just add and create the relationship between it and the node needed. And when you want to expand your reach to other areas, we just need to add a new Node Place

and link to the old place is easy. Figure 3.2.8 is an example when we extend to the Other cities, the management of the City is similar to Place. If we expand to another country, we do the same thing.

This makes it easy to manipulate data and handle large amounts of data. GraphDB has brought benefits to data processing and this is why we chose GraphDB.

3. New Feature: Self-made tour

Users can create Tripbooks - travel created by themself - by pressing the “+” button to add spots or events they like on Tripbook. They can change their order, location, route, comment and comment on their tripbook. When locations in Tripbook have been prepared they can connect the locations.

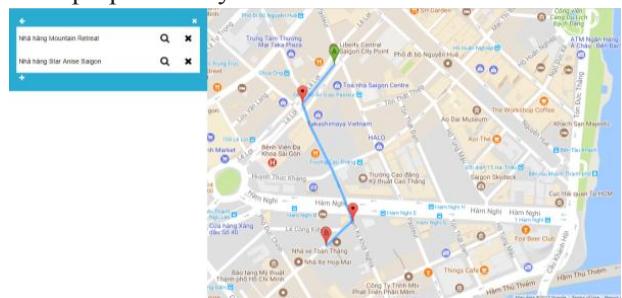


Figure 3.3.1: Mapping place

Users can contribute back to the system by sharing. They can evaluate locations, events by tapping, commenting, star rating. When satisfied with his/her Tripbook users can share it with everyone. In this way, people will have more information to choose places and plan their trips, as well as providers of places and events, can know where people are judged on locations or locations they providing. Then they can change to make them better.

IV. DISCUSSION

1. Features's Comparision

Foody and CocCoc Map are two of the most prominent

location-based search applications (restaurants, restaurants, hotels) in Vietnam with massive data and continuous updates. CocCoc Map provides users with a detailed map search feature with an intuitive map view, but the users only stops at the search level. With Foody, they have a built-in list interface that is designed to meet the user experience, direct links to locations, and adding user reviews that make their applications more user-friendly and reliable but they have not yet provided users with a private trip creation feature. Pollia will focus on this unique trip creation feature to make it different, and we're also working to improve performance gradually so that we can meet our users' needs.

2. Database Query Comparison

For sales applications, food information, or other types of information and tend to expand without changing the structure, Graph Database optimizes this. The speed with which you access data and how to call Query is also easier.

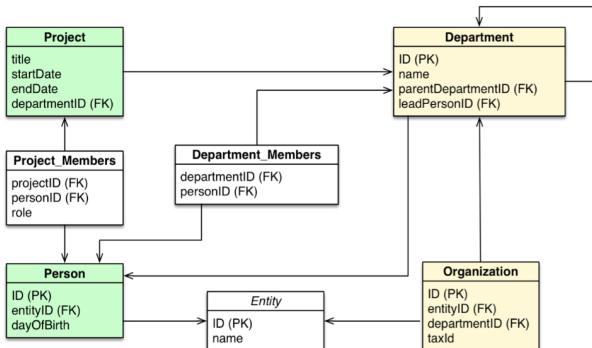


Figure 4.2.1: Example of RDBMS data organization

```

SELECT product.product_name as
Recommendation, count(1) as Frequency
FROM product, customer_product_mapping,
(SELECT cpm3.product_id, cpm3.customer_id
FROM Customer_product_mapping cpm, Customer_product_mapping cpm2, Customer_product_mapping cpm3
WHERE cpm.customer_id = 'customer-one'
and cpm.product_id = cpm2.product_id
and cpm2.customer_id != 'customer-one'
and cpm3.customer_id = cpm2.customer_id
and cpm3.product_id not in (select
distinct product_id
FROM Customer_product_mapping cpm
WHERE cpm.customer_id = 'customer-one')
) recommended_products
WHERE customer_product_mapping.product_id =
product.product_id
and customer_product_mapping.product_id in
recommended_products.product_id
and customer_product_mapping.customer_id =
recommended_products.customer_id
GROUP BY product.product_name
ORDER BY Frequency desc
    
```

Figure 4.2.2: SQL Query

```

MATCH (u:Customer {customer_id:'customer-one'}) -[:BOUGHT]->(p:Product)-<-
[:BOUGHT] - (peer:Customer) -[:BOUGHT]->
(reco:Product)
WHERE not (u) -[:BOUGHT]->(reco)
RETURN reco as Recommendation, count(*)
as Frequency
ORDER BY Frequency DESC LIMIT 5;
    
```

Figure 4.2.3: Cypher Query

Figure 4.2.1 is an example of how we compare two query statements in two query languages, SQL for RDBMS and the GraphDB Cypher. The meaning of the statement we will use is: for each customer to buy the same product, find the product that the customer left to buy that customer does not buy.

We can see in Figure 4.2.2 and 4.2.3 that while Cypher is written three times faster than SQL Query. There are no differences in performance due to the JOIN statement, but it also reduces performance when writing query statements. With Cypher, we can more easily understand how to access data than SQL, writing shorter leads to more time spent on other things. And we also test the performance when using RDBMS and GraphDatabase by Neo4j. The result is good like we expect.

V. CONCLUSION AND FUTURE WORK

In this paper, we have discussed the need for a technology solution to meet the travel trends of people especially young people today. As well as looking at the applications currently available in the market for mapping and locating locations in Vietnam, those applications still do not meet the needs mentioned above. So, we have researched and created Pollia as a software solution that allows users to create their own trip to experience the self-guided travel experiences. To enrich the data sources on information, forms of tourism and culture of places and events in Vietnam; Help providers provide an overview of user reviews to improve the destinations they provide. In terms of data, we have used NoSQL and, in particular, Neo4j is a type of GraphDB that integrates with SQL data to improve query speed, which improves system performance for a better experience for users.

In the future, we will strive to build relationships with sites to increase application credibility and dissemination to site owners or service providers. Build a system to calculate the average cost and average time spent by a person visiting a place and event. Build a tourism community where people seek to share experiences about places, events and trips. People can exchange for the development of the tourism community, thereby indirectly contributing to the process of forming the database. This results in

more and more data leading to the fact that we have to apply Bigdata to manage and exploit effectively in the future.

ACKNOWLEDGMENT

We would like to thank the teachers at FPT University for teaching us the important foundation of software software industry. We would like to thank our colleagues, friends and others who have helped us a lot in terms of experience, contributing ideas in completing the research. This paper can not be completed without a dedicated direction and always follow us in the process of implementation of Huynh Cong Viet Ngu (FPT University lecturer).

REFERENCES

- [1] Mehdi Jazayeri faculty of Informatics Faculty of Informatics University of Lugano (USI) Technical University of Vienna Lugano, Switzerland Vienna, Austria, "Some trends in Web Application"
- [2] Pramod J. Sadalage | Martin Fowler "NoSQL Distilled"
- [3] Philippe Rigaux, Michel Scholl, and Agnes Voisard. Morgan Kaufmann, 2002, "Spatial Databases with Application to GIS"
- [4] Michael Hunger, Ryan Boyd, William Lyon, "RDBMS & Graphs: SQL vs. Cypher Query Languages"
- [5] Huỳnh Công Việt Ngữ Department of IT FPT University HCMC Campus, "A Ship Performance Monitoring System Supporting Remote Monitoring (Korea Multimedia Society Spring Conference 2016, Seoul)"
- [6] Huỳnh Công Việt Ngữ Department of IT FPT University HCMC Campus, "Construction of an R-tree using STR Partitioning and MapReduce (Korea Multimedia Society Autumn Conference 2016, Busan)"
- [7] Huỳnh Công Việt Ngữ Department of IT FPT University HCMC Campus, "B-tree Construction with Huge volume of Data on Hadoop (SCECR – the Statistica Conference in E-Commerce Research - Original Big data, Economic &Law University, under Vietnam National University-Ho Chi Minh city)" • Journals:
- [8] Huỳnh Công Việt Ngữ Department of IT FPT University HCMC Campus, "Improving the Quality of an R-tree Using the MapReduce Framework (International Conference on Ubiquitous Information Technologies and Applications (CUTE 2016), 19-21/12/2016, Bangkok, Thailand, Advanced Multimedia and Ubiquitous Engineering, (CUTE 2016), vol. 448, pp. 164–170. Springer, Singapore (2017) – recommend to SCI)"

[9] Huỳnh Công Việt Ngữ Department of IT FPT University HCMC Campus, "Improving the B+-Tree Construction for Transaction Log Data in Bank System Using Hadoop (iCatse Conference on Information Science and Applications (ICISA 2017), 20-23/3/2017, Macau, China, was published on LNEE424, ICISA2017, Book: Information and Applications 2017, Chapter No:60, Chapter DOI:10.1007/978-981-10-4154-9_60, Page 519-525- recommended to SCI journal (Cluster Computing))"

[10] Huỳnh Công Việt Ngữ Department of IT FPT University HCMC Campus, "B+tree construction on Massive data with Hadoop (Cluster Computing – DOI 10.1007_s10586-017-1183-y – recommended from ICISA 2017)"

[11] Đặng Hoàng Minh - Từ hái rau thuê tới sáng lập Foody at "http://pace.edu.vn/so-tay-doanh-tri/ChiTiet/1089/dang-hoang-minh-tu-hai-rau-thue-toi-sang-lap-foody?term_taxonomy_id=31"

[12] Cốc Cốc ra mắt bản đồ cạnh tranh Google Maps: Chi tiết đến từng quán trà đá, tiệm bom xe at "<http://genk.vn/coc-coc-ra-mat-ban-do-canhanh-google-maps-chi-tiet-den-tung-quan-tra-da-tiem-bom-xe-20171123141709328.chn>"

Papers of Technical Sector

A Framework for Clustering Time Series Data

Vuong Chanh Quyen
FPT University
HCMC, Viet Nam

Trinh Thanh Quyen
FPT University
HCMC, Viet Nam

Nguyen Minh Nhut
FPT University
HCMC, Viet Nam

Abstract

Time series data play an important role in various fields from science, technology, medicine to economics such as user data management, business planning, meteorological forecasting, etc. In fact, time series data are often very large and varied, causing difficulties in managing and analyzing them. Therefore, clustering time series data is essential to classify time series and develop Decision Support Systems (DSS). However, there is no product available in the market to assist users in clustering time series data efficiently. So, a framework, running on Windows, is designed and implemented for clustering time series data. In this document, information related to time series data will be introduced including algorithms for clustering data, methods for evaluating clustering results.

Keywords

Data clustering; Time series data; Framework; k-means algorithm; Evaluation method; Data visualization

I. INTRODUCTION

Time series is a sequence of data points, measured in successive moments of time in a uniform time frame [8]. In our life, time series is a common type of dynamic data that naturally arises in many different scenarios, such as economic data, forecast data, electrocardiogram, etc.

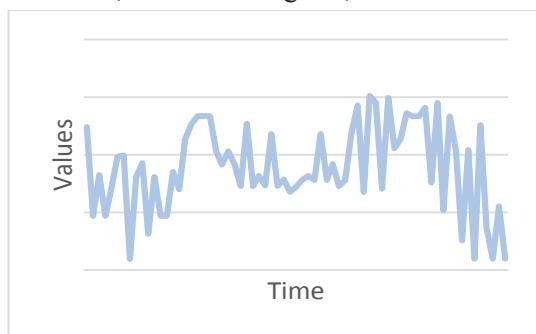


Figure 1: Time series data graph of earthquake

Supervisor: Than Van Su, M.E (Head of Academic Affairs and Lecturer of IT Department, HCMC FPT University)
Defended topic in front of the scientific council of HCMC FPT University on December 13rd, 2017
Accepted post on December 26th, 2017

Basically, Time series data fall into two types:

Type 1: Series had been clustered and stored in knowledge base, and they are used in DSS, such as:

- Sick people's electrocardiogram, normal people's electrocardiogram had been clustered by doctor...
- The activities of nature such as earthquake, tide, tsunami, sound, etc. are recorded for years.

If there is new concept that is introduced, clusters need to be re-clustered.

Type 2: Data have not been clustered yet, that need to be clustered.

When work with time series, pre-processing of data must be done [3]:

- Length and scale normalization.
- Characteristic line definition.

Clustering data is the task of grouping a set of data objects into data clusters so that objects in the same cluster are more similar to each other than objects in the other clusters. Clustering time series data is make it into clusters in order to having an overview of the differences of each cluster as well as the nature of them [2].

In this paper, the primary goal is to introduce a framework for clustering time series and to provide an overview of time series clustering result.

The remainder of this paper is organized as follow. Section II introduces the problem and the solution. The next section describes the structure of framework and the used algorithms. And in the final section, experimental results as well as conclusion are presented.

II. PROBLEM & SOLUTION PLAN

Nowadays, with the incremental development of data storage and retrieval, data clustering, especially time series, has been shown effective in providing useful information in various domains [9]. Because of the needs of clustering data and reusable application, the framework must be created. The framework has DLL files including some required methods such as: load elements data, cluster data, evaluate accuracy of algorithms, visualize time series, etc.

In fact, time series data are often varied and has large size. So calculating the distance between two time series as well as clustering the time series using the Euclidean distance takes longer than using the square

of the Euclidean distance. To cluster time series quickly and efficiently, the framework is designed and developed based on K-Means algorithm and square of Euclidean distance.

III. PLAN IMPLEMENTATION

A. Framework's Structure

The structure of framework includes the following interfaces and default classes:

Interfaces:

I_Element: prototypes methods of an element

Table 1: I_Element interface detail

Methods	Purpose	Input	Output
Get Euclidean Distance	Compute distance between two elements.	Element.	Distance between two elements.
GetSquare Euclidean Distance	Compute square distance between two elements.	Element.	Square of distance between two elements.
LoadData	Loading data from file. Double values is required.	Path of file.	Data are loaded
Draw Element	Draw element on the coordinate axis.	Drawing area is identified by position, width and height.	Element is drawn on drawing area in coordinates axis.

I_Cluster: prototypes methods of a cluster

Table 2: I_Cluster interface detail

Methods	Purpose	Input	Output
Setup	Setup data for cluster.	List of elements.	List of elements is set to cluster's element list.
Load Data	Load data from file.	Array of file's path.	Data are loaded into element list.

Write Result	Export cluster and elements to folder and file.	Cluster name.	Cluster is exported to folder and files.
Get Element	Get element from cluster	Index of element.	Element at index.
Add Element	Add element to cluster	Element	Result when add element.
Draw Cluster	Draw all elements of cluster on coordinate axis.	Drawing area is defined by position, width and height.	All cluster's elements are drawn on drawing area in coordinates axis.

I_Evaluator: prototypes method for evaluation

Table 3: I_Evaluator interface detail

Methods	Purpose	Input	Output
Evaluate	The results of cluster can be evaluated base on evaluation coefficients	List of elements, list of result clusters and list of clusters to evaluate.	Evaluation results

Default Classes:

DefaultElement: implements methods prototyped in the I_Element interface

Table 4: DefaultElement attributes	
Data Members	Description
Values	Values of element that must be set of double numbers.
Name	Element's filename.

Table 5: DefaultElement methods

Individual Methods	Purpose	Input	Output
Plus	Plus values of two elements.	Default	

Divide	Divide element's values by an integer.	Element	Result element that have values is the sum of two element's values.
Equals	Compare two element	Integer	Result element that have values are quotient.
		Element	Result of comparing two elements

DefaultCluster: implements methods prototyped in the *I_Cluster* interface

Table 6: DefaultCluster attributes

Data Members	Description
Center	Center of cluster.
ElementList	List of cluster's elements.

Table 7: DefaultCluster methods

Individual Methods	Purpose	Input	Output
GetMax Value	Get maximum value among elements of cluster.		Maximum value
GetMin Value	Get minimum value among elements of cluster.		Minimum value

KMeans: implements methods prototyped in the *I_Evaluator* interface

Table 8: KMeans methods

Individual Methods	Purpose	Input	Output
ClusterData	Cluster elements into clusters.	Number of clusters (k) and elements.	List of result clusters.

B. Algorithms

• K-Means Algorithm

K-means algorithm is one of the oldest and most commonly used clustering algorithms. The K-means clustering algorithm is described in detail by Hartigan (1975). The aim of the K-means algorithm is to

divide M points in N dimensions into K clusters so that the within-cluster sum of squares is minimized [4].

The K-Means algorithm is described as below:

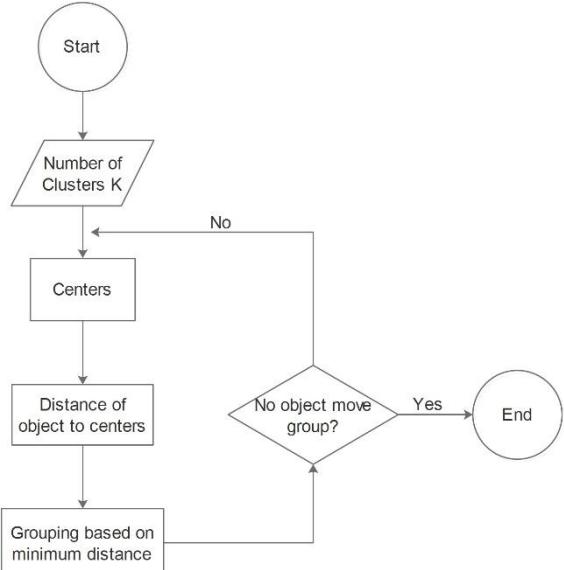


Figure 2: K-Means algorithm flowchart

With n elements $E = \{E_0, \dots, E_{n-1}\}$, the K-Means algorithm performs the following steps:

- Step 1: Select k random centers $C = \{C_0, \dots, C_{k-1}\}$. Each cluster is represented by the center of the cluster.
- Step 2: For each element, calculate the square of Euclidean distance between element to K centers. The formula for square of Euclidean distance between a point X (x_1, x_2, \dots) and a point Y (y_1, y_2, \dots) is:

$$D^2 = \sum_{i=0}^n (x_i + y_i)^2$$

- Step 3: Group the elements into the nearest group.
- Step 4: Recalculate the center for groups.

The formula for center C of cluster with n elements $\{E_1(x_1, y_1, \dots), E_2(x_2, y_2, \dots), \dots\}$ is:

$$C = \left(\frac{x_1 + x_2 + \dots + x_n}{n}, \frac{y_1 + y_2 + \dots + y_n}{n}, \dots \right)$$

Repeat step 2 until there are no changes of objects in groups.

The result of K-Means algorithm is a collection of k clusters $\{C_1 = \{E_1, E_2, \dots\}, \dots, C_k = \{E_1, E_2, \dots\}\}$.

• Balance centers algorithm

Because the centers initiation of the K-Means algorithm is random, in some cases the algorithm will give inaccurate results. In order to increase the accuracy of the K-Means algorithm, the initial centers should be chosen by a systematic method [5]. With n elements and k random centers, the followings

can be calculated:

- C_n : center of n elements.
- C_k : center of k centers.
- D: distance between C_k and C_n .

By minimizing the distance (D) between the center of k centers (C_k) and the center of n elements (C_n), the balance centers algorithm adjusts the centers so that they are equally distributed and achieve a certain balance. Then the K-Means algorithm will be executed faster and more accurately.

The balance centers algorithm is described as below:

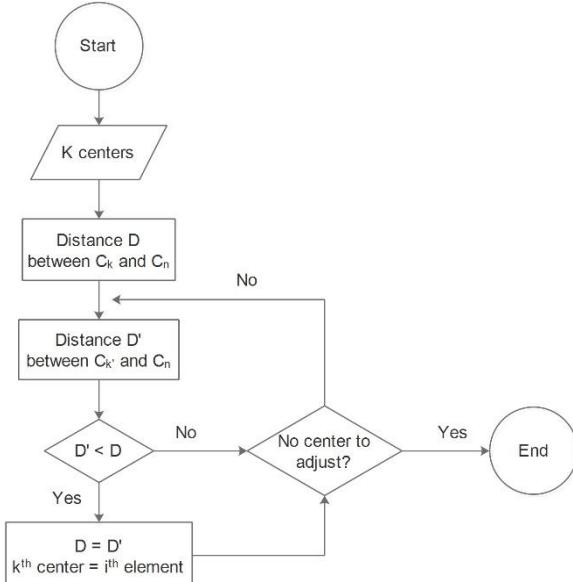


Figure 3: Balance centers algorithm flowchart

With n elements $E = \{E_0, \dots, E_{n-1}\}$, the balance centers algorithm performs the following steps:

- Step 1: Select k random centers $C = \{C_0, \dots, C_{k-1}\}$.
- Step 2: Calculate center of k centers (C_k), center of n elements (C_n), and square of Euclidean distance (D) between C_k and C_n .

The formula for center C of n elements $\{E_1(x_1, y_1, \dots), E_2(x_2, y_2, \dots), \dots\}$ is:

$$C = \left(\frac{x_1 + x_2 + \dots + x_n}{n}, \frac{y_1 + y_2 + \dots + y_n}{n}, \dots \right)$$

The formula for square of Euclidean distance between a point X (x_1, x_2, \dots) and a point Y (y_1, y_2, \dots) is:

$$D^2 = \sum_{i=0}^n (x_i + y_i)^2$$

- Step 3: for each element, replace kth center ($c[k]$) with ith element ($e[i]$) then recalculate center (C_k') and distance (D'). If $D' < D$, then $c[k] = e[i]$ and $D_1 = D_2$.

- The adjusted centers will be marked to prevent them from being selected again.

When the algorithm stops, all centers are adjusted.

The result of Balance centers algorithm is a collection of k adjusted centers $C = \{C_1, \dots, C_k\}$.

C. Evaluation Coefficients

To identify the accuracy of algorithm, the most popular clustering evaluation method is comparing a clustered result with a classified dataset. Five evaluation coefficients are used in application: Jaccard, Rand, FM and CSM used for evaluating time series clustering algorithms, and NMI used for validating clustering results [11].

Consider $G = G_1, G_2, \dots, G_M$ as the clusters from a supervised dataset, and $A = A_1, A_2, \dots, A_M$ as that obtained by a clustering algorithm under evaluations. Denote D as a dataset of original time series or features. For all the pairs of series (Di, Dj) in D, we count the following quantities:

- a is the number of pairs, each belongs to one cluster in G and are clustered together in A.
- b is the number of pairs that are belong to one cluster in G, but are not clustered together in A.
- c is the number of pairs that are clustered together in A, but are not belong to one cluster in G.
- d is the number of pairs, each neither clustered together in A, nor belongs to the same cluster in G.

The used clustering evaluation coefficients are defined as below:

1. Jaccard Score (Jaccard):

$$Jaccard = \frac{a}{a + b + c}$$

2. Rand statistic (Rand):

$$Rand = \frac{a + d}{a + b + c + d}$$

3. Fowlkes and Mallows index (FM):

$$FM = \sqrt{\frac{a}{a+b} \times \frac{a}{a+c}}$$

4. Cluster Similarity Measure (CSM):

$$CSM(G, A) = \frac{1}{k} \sum_{i=1}^k \max_{1 \leq j \leq k} Sim(G_i, A_j)$$

where: $Sim(G_i, A_j) = \frac{2|G_i \cap A_j|}{|G_i| + |A_j|}$

5. Normalized Mutual Information (NMI):

$$NMI = \frac{\sum_{i=1}^k \sum_{j=1}^k n_{i,j} \log \left(\frac{n \cdot n_{i,j}}{|G_i| |A_j|} \right)}{\sqrt{\left(\sum_{i=1}^k |G_i| \log \frac{|G_i|}{n} \right) \left(\sum_{j=1}^k |A_j| \log \frac{|A_j|}{n} \right)}}$$

where: $n_{i,j} = |G_i \cap A_j|$

All used coefficients have value ranging from 0 to 1, where 1 corresponds to the case when G and A are identical.

IV. EXPERIMENTAL RESULTS & CONCLUSION

To demonstrate framework's features, an application, running on Windows, is designed in Microsoft Visual C# 2010 and uses .NET Framework 4.5. The computer system uses CPU Intel Core i3-2350M, 6GB RAM, and runs OS Windows 10 Pro.

The given set of experimental data contains 500 time series data including burst, earthquake, tide, etc. (Figure 4). Size of each time series data is 1024. Time series in the given set have different basic shape and are classified into 10 clusters. Time series data in the same cluster have the same characteristic line and the data amplitude equivalent to each other. In addition, time series data in the same cluster will have the same name, differentiated to each other by the code in file name, such as burst_1, burst_2, etc.

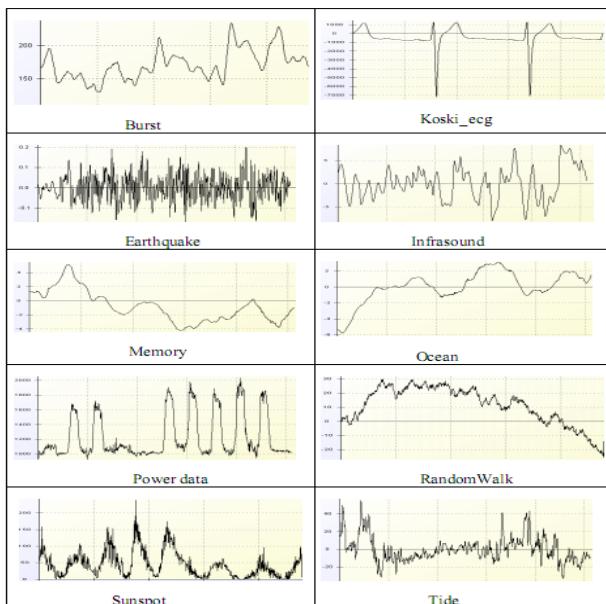


Figure 4: Experimental dataset

Evaluation of the framework will be done by evaluating quality of K-Means algorithm in clustering data.

• Clustered dataset

The evaluation will be done with the incrementally number of elements (100, 300, 500) to test the accuracy as well as the execution time of algorithm.

Table 9: Evaluation results

Size	Jac-card	Rand	FM	CSM	NMI
100	1.0	1.0	1.0	1.0	1.0
300	1.0	1.0	1.0	1.0	1.0
500	1.0	1.0	1.0	1.0	1.0

Table 10: Execution time results

Size	100	300	500
Execution Time	774 ms	2564 ms	4120 ms

Based on the results of the evaluation, we can see that the algorithm achieves high accuracy in the measurements. In addition, the execution time of the algorithm increases as the amount of data increases. In general, with the execution time as above, the application of the framework for clustering data in large systems is possible.

• Non-clustered dataset

With time series data that were not clustered, the framework supports data visualization to help users determine the accuracy of clustering data based on the characteristic line. If data clustering is correct, the visualized data will have the same characteristic line (figure 5). Otherwise, the visualized data will have different characteristic lines (figure 6) [10].

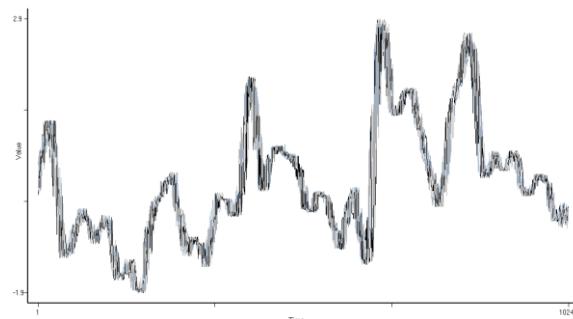


Figure 5: Data with same characteristic line

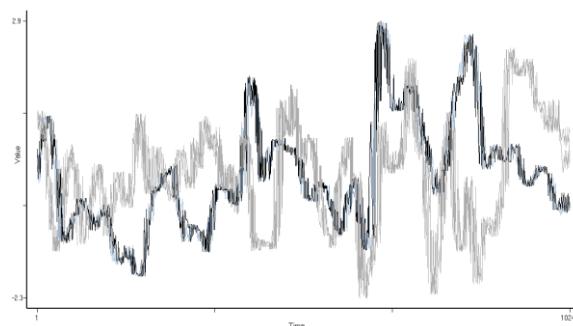


Figure 6: Data with different characteristic lines

Based on the experimental results, the following conclusions can be drawn:

• Advantages

- The algorithms implemented in the framework have high accuracy and acceptable execution time, so it can be applied in time series clustering. Clustered data will be used in decision support systems and data warehouses.

- The framework also supports data visualization to help users cluster new and unclassified data.

• Disadvantages

- The framework does not yet support data normalization. Before clustering data, users must spend time normalizing data into a uniform.
- The framework currently only supports clustering data with the k-means algorithm, so it is not possible to cluster time series data having high noise and complexity.

In the future, the framework will be developed to cluster many types of data using various clustering algorithms.

REFERENCES

- [1] Alsabti, K., Ranka, S., and Singh, V., 1998, An Efficient K-Means Clustering Algorithm. In Proceedings of IPPS/SPDP Workshop on High Performance Data Mining.
- [2] Anil K. Jain, Data Clustering: 50 Years Beyond K-Means, Pattern Recognition Letters, Vol. 31, No. 8, pp. 651-666, 2010.
- [3] E. Ogasawara, L. C. Martinez, D. de Oliveira, G. Zimbrao, G. L. Pappa, and M. Mattoso, Adaptive Normalization: A Novel Data Normalization Approach for Non-Stationary Time series, International Joint Conference on Neural Network (IJCNN), IEEE, 2010.
- [4] J. A. Hartigan and M. A. Wong, Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 28, No. 1 (1979), pp. 100-108.
- [5] Pena, J., Lozano, J., and Larraaga, P., 1999, An Empirical comparison of four initialization methods for the k-means algorithm. In Pattern Recognition Letters Vol. 20, pp. 1027-1040.
- [6] Spencer Schnier. Choosing Distance Measures for Mining Time Series Data. University of Illinois at Urbana Champaign, USA, February 2011.
- [7] Tatsunokuchi, Ishikawa, Minyang, Sichuan (2004), An evolutionary K-means algorithm for clustering time series data. In: 3rd International Conference on Machine Learning and Cybernetics. Shanghai, pp 1282–1287.
- [8] Time series in Wikipedia. [https://en.wikipedia.org/wiki/Time_series]
- [9] Warren Liao, T., 2005, Clustering of time series data—a survey. In Pattern Recognition, Volume 38, Issue 11 (November 2005), pp. 1857-1874.
- [10] Weber, M., Alexa, M., and Muller, W., 2001, Visualizing Time Series on Spirals. In Proceedings of 2001 IEEE Symposium on Information Visualization. San Diego, CA, Oct 21-26. pp. 7-14.
- [11] Zhang, Ho, T. B., Zhang, Y., Lin, M. S., 2006, Unsupervised Feature Extraction for Time Series Clustering Using Orthogonal Wavelet Transform. Journal of Informatica, Vol. 30, No. 3, 2006, pp. 305-319.

Applying iBeacon Technology to build Automation Tollbooth Station:

Combined with vehicle number plate recognition technique

Le Vuong Quoc Huy
FPT University
HCMC, Viet Nam

Nguyen Chi Hieu
FPT University
HCMC, Viet Nam

Pham Bao Tin
FPT University
HCMC, Viet Nam

Do Viet Son
FPT University
HCMC, Viet Nam

Abstract

One problem in traffic sector is to build non-stop toll station. The difficulty in building non-stop toll station is how to detect when a vehicle is coming and, which vehicle it is and how to notify for drivers to make payment. The solution currently applied to build automation tollbooth is using RFID technology, belongs to VETC Corp. This solution still has some limits such as the inability to control the range of signal as well as the instability of its signal. Using bluetooth low energy signal, iBeacon technology provide the improvement in notifying drivers when their vehicles passing toll-booth station. Besides, without installation the RFID tag into vehicle, it will be simpler for drivers to register to system and they can control the payment process through mobile application in smart phone. The number plate recognition supports the system to capture the evidence which vehicle passing the station in real time.

Keywords

iBeacon; BLE (Bluetooth Low Energy) signal;
RFID; License Plate Recognition

I. INTRODUCTION

Building the automation toll-booth system is an important part of improving the traffic situation in Viet Nam. The non-stop toll station could bring benefits such as reducing traffic jams and pollution, saving energy and time for drivers, and saving investment cost. Due to its significance, VETC Company provided the solution using RFID Technology in identifying vehicle passing the toll station. When a driver or vehicle's owner wants to register to system, they have to go to VETC Center and stick a VETC tag into their vehicle as well as open a VETC account.

This tag must be kept intact, not disassembled. When vehicle goes to toll station implemented VETC System, it must slow down. At the station, there are some machines broadcast the radio frequency, the tag in vehicle receives the frequency and accumulates into energy to response the information of vehicle. Then, the system identifies the which vehicle is passing to process payment and keeps track its data. This solution has been approved and implemented by the government. According to the plan, in 2020, all toll stations will apply this technology.

In this paper, we approach how to detect vehicle in a new way. Driver will be the one identifying when they come to a toll station and decide to make payment. Besides, at each station, camera system with license plate recognition supports capturing the evidence whenever a vehicle passing.

Beacons with identification information installed at each station will broadcast bluetooth signal, which is more stable and easier to cast than radio frequency. Drivers use mobile application to receive bluetooth signal to know when they go inside a toll station. Then, they are able to be actively agree or reject to pay the fee. To prevent violation, the camera system will monitor the toll station 24/7 to detect when a vehicle comes and what is its license plate. When the license plate is detected, it is easy to keep track the vehicle's data and handle the violation.

The primary advantage of our solution we feel is providing the convenient and proactive method for driver to pay the traffic fee. Drivers could pay the fee automatically or manually or later because their vehicle had been captured when they pass the toll station.

In section 2, we formalize the problem and describe our approach to solve it. In section 3, we present the database design for configuring beacons and algorithms for detecting and recognizing license plate of vehicle. We analyze the solution in section 4. Experimental results and the software usage is presented in section 5.

II. PROBLEM AND SOLUTION PLAN

1. How to detect the toll station

For building the automation toll station in this new

solution, we need a way to make mobile app used by driver distinguish when it enters or exits a toll station and which toll station. Therefore, the first necessary condition for mobile app is ability to check which toll-station it is staying inside and how to get the information of that station.

This condition leads to the need of tracking location of mobile app. First thought of building a location-based application is using GPS; however, in this case the high exactness is required, moreover mobile app must operates immediately when it enters the region of toll station.

To solve this problem, we use the approach of using iBeacon devices provided by Apple. The iBeacon technology allows Mobile Apps to understand their position on a micro-local scale, and deliver hyper-contextual content to users based on location. The underlying communication technology is Bluetooth Low Energy (BLE). BLE communication consists primarily of small packets of data, broadcast at a regular interval by Beacons or other BLE enabled devices via radio waves. Apple has standardized the format of those packets of data. Under this format, an advertising packet consists of four main pieces of information: UUID, Major, Minor and Tx Power fields. First three field is used to map each beacon with one specific object, such as one lane in one toll station. The last field is used to determine the distance between mobile apps and beacon device.

2. How to recognize license plate by camera

By using iBeacon, it might be not enough for preventing violation. We must provide a method for capture the evidence when a vehicle passes the toll station. Therefore, we build a license plate recognition technique together.

We use the camera system to monitor all lanes at each toll station. The difficulty we have to solve is how to detect when a vehicle goes inside to capture the photo and how to recognize the license plate from that photo.

Currently, we just begin to research the technique of image recognition and develop the method how to detect text and number in an image. The solution up to now is controlling the camera to monitor the road. Then, we stream data from camera as many frames of photo. With each frame, we apply the algorithms to find the suitable license plate. If there is any license plate found, we validate that plate and send to system center to process and storage.

In next section, we will describe in detail about how to set up data for beacon mapping as well as the algorithms to recognize text in image.

III. PLAN IMPLEMENTATION

1. Apply beacon technology

1.1. Configure beacon mapping

Beacons that want to be “discovered” can broadcast packets of data in set intervals. These packets are meant to be collected by devices like smartphones, where they can be used to trigger things like push messages, app actions, and prompts (Figure 1.). As we mentioned, a small packet of data broadcast by each beacon includes UUID, Major, Minor are used to identify and distinguish between beacons.

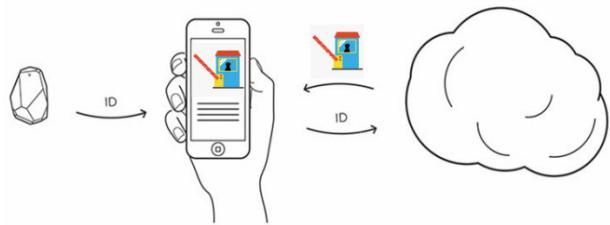


Figure 1

UUID: This is a 16 byte string used to differentiate a large group of related beacons.

Major: This is a 2 byte positive integer number used to distinguish a smaller subset of beacons within the larger group.

Minor: This is a 2 byte string meant to identify individual beacons.

To establish a beacon network for this system, we choose a specific UUID for all beacon. This UUID helps mobile app know that it comes to region of beacons belong to automation toll-station system but no other beacons. Here we choose the 16-byte string B9407F30-F5F8-466E-AFF9-25556B572896.

Next, the major field help to distinguish a group of beacons within our beacon network. In our system, at each station, we need a group of beacons operates different functions. Therefore, major field is used to determine the beacon belongs to which toll station. For example, we choose number 24100 for toll station Cai Lậy in Tiền Giang province.

Next step, at each toll station, we need each beacon installed in each lane to monitor which vehicle runs through the toll station on which lane. Hence, each beacon at each lane will be configured with one minor number.

For example, a beacon broadcasts the following packet (B9407F30-F5F8-466E-AFF9-25556B572896:241 00:100). The device receiving this packet would understand it's from the beacon belongs to lane [100] of toll station Cai Lậy, Tiền Giang province which has been installed this system. Major and Minor field is unsigned 2-byte integer number, so it could be up to 65535 number. It means we could have more than 60 thousand station within this beacon network.

1.2. Set up event handler on mobile apps

Estimote Beacon provide API for android and iOS smartphone. According to this API, we could set up handler for event “enter” or “exit” a beacon region. In iBeacon, beacon region can be defined in three ways:

With only UUID: it consists of all beacons with a given UUID. For example: a region defined with default Estimote UUID would consist of all Estimote Beacons with unchanged UUID.

With UUID and Major: it consists of all beacons using a specific combination of UUID and Major. For example, all Estimote Beacons with default UUID and Major set to 13579.

With UUID, Major and Minor: it consists of only a single beacon (Estimote Cloud prevents having two beacons with the same IDs). For example, one with default Estimote UUID, Major set to 13579 and Minor set to 2468.

To enhance performance, our mobile app just monitor the beacon region with UUID only. It means mobile app will be triggered whenever it enters any region of beacon has UUID B9407F30-F5F8-466E-AFF9-25556B572896 but no other UUID.

When device enters beacon region, it receives the packet from that beacon. Then, we analyze the packet to get the UUID, Major and Minor fields. We could use those data to get information of toll station device is running into. After that, we let the user (driver) to decide to pay the fee for that toll station or not.

Keep using the example above, when device enters the region and receives the packet, it knows that is lane [100] of Cai Lậy toll station. It also could get the price of this station. Then, there is a prompt for driver to agree or reject to pay the fee.

Same as “enter” event, “exit” event will be triggered when device gets out of region of any beacon has UUID above. In “exit” event, we will request server to check the transaction result in case driver agreed to pay fee before.

1.3. Map beacon packet with toll station information

Next, we need a process to map beacon packet to toll station’s information in order to help mobile app understand where it is. This purpose leads to the requirement of data-base designing.

We need to build a list of toll stations which installed this system. Each record includes information of that toll station, includes its name, its location, its city/province, the list of price corresponding to each vehicle type. Following is the example table of toll station data:

	id	name	location	zone	isActive
1	Cai Lậy	[10.412456, 106.064289]	Tiền Giang	1	
2	Gò Dầu	[11.140563, 106.276213]	Tây Ninh	1	
3	Đèo Bả...	[11.477144, 107.737985]	Lâm Đồng	1	

Then, in each table, we need 2 or 3 lanes, each controlled by a beacon device:

id	name	stationId	isActive
1	Lane 1	1	1
2	Lane 2	1	1
3	Lane 1	2	1
4	Lane 2	2	1
5	Lane 3	2	1
6	Lane 1	3	1
7	Lane 2	3	1

Finally, we have a table of beacons, each beacon has three field UUID, Major and Minor to identify and id of lane and id of station for mapping:

id	uuid	major	minor	stationId	laneId
1	B9407F30-F5F8-466E-AFF9-25556B572896	36857	31381	1	1
2	B9407F30-F5F8-466E-AFF9-25556B572896	36857	3245	1	2
3	B9407F30-F5F8-466E-AFF9-25556B572896	39748	38452	2	1
4	B9407F30-F5F8-466E-AFF9-25556B572896	39748	5534	2	2
5	B9407F30-F5F8-466E-AFF9-25556B572896	39748	3434	2	3
6	B9407F30-F5F8-466E-AFF9-25556B572896	24100	4804	3	2
7	B9407F30-F5F8-466E-AFF9-25556B572896	24100	7578	3	2

Finally, whenever the auto handler receives a request from mobile app, it could analyze to get three field UUID, Major and Minor. Then, it queries database to get the id of station and id of lane corresponding with the beacon packet. Then, it gets the information of station and lane from two table above and responses to mobile app. Mobile app receives response and understand the toll station it is coming inside.

2. Apply license plate recognition algorithms

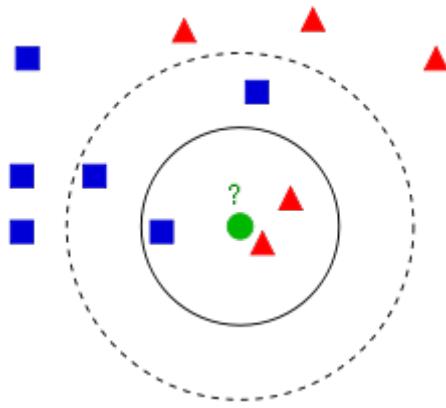
2.1. K-Nearest Neighbor Algorithms

K-Nearest Neighbor (KNN) algorithms is usually used for pattern recognition problems, here is detecting text in image. It is commonly used for its easy of interpretation and low calculation time. KNN Algorithms has two phases, training and classification.

In training phase, we define the set of “neighbors”. The neighbors are taken from a set of objects for which the class (for k-NN classification) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples.

In the classification phase, k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point.

A commonly used distance metric for continuous variables is Euclidean distance. For discrete variables, such as for text classification, another metric can be used, such as the overlap metric (or Hamming distance). In the context of gene expression microarray data, for example, k-NN has also been employed with correlation coefficients such as Pearson and Spearman. Often, the classification accuracy of k-NN can be improved significantly if the distance metric is learned with specialized algorithms such as Large Margin Nearest Neighbor or Neighborhood components analysis.



Example of k-NN classification. The test sample (green circle) should be classified either to the first class of blue squares or to the second class of red triangles. If $k = 3$ (solid line circle) it is assigned to the second class because there are 2 triangles and only 1 square inside the inner circle. If $k = 5$ (dashed line circle) it is assigned to the first class (3 squares vs. 2 triangles inside the outer circle).

The best choice of k depends upon the data; generally, larger values of k reduce the effect of noise on the classification, but make boundaries between classes less distinct. A good k can be selected by various heuristic techniques (see hyperparameter optimization). The special case where the class is predicted to be the class of the closest training sample (i.e. when $k = 1$) is called the near-est neighbor algorithm.

Next, we will present the steps of applying KNN Algorithms in recognition the license plate in each frame streamed from camera.

2.2. Preprocessing Image

Firstly, we streams images recorded by camera frame by frame. With each image, we represent it as three matrices of number. The size of matrices is the size

of image. Each matrix specify the amount of Red, Green Blue that makes up the image. The number in these matrices is integer number between 0 and 255, and they determine the intensity of the pixel with respect to the color of the matrix.

Step one, we convert an original image to gray scale image. As we mentioned above, each pixel in image is a combination of three number corresponding to amount of red, green, blue color make up that color pixel. The grayscale algorithms use three-step process: get the red, green, blue values of each pixel; turn those number into a single gray value; replace the original red, green, blue values by new gray value. The formula we use to turn color values into gray value is:

$$\text{Gray} = (\text{Red} * 0.299 + \text{Green} * 0.587 + \text{Blue} * 0.114)$$

This formula is used by some modern digital image and video for slightly different coefficients.

Step two, we need to thresh hold the grayscale image. This process is used for create binary image from grayscale image. The simplest thresholding methods replace each pixel in an image with a black pixel (min value) if the image intensity $I_{i,j}$ is less than some fixed constant T (that is, $I_{i,j} < T$), or a white pixel (max value) if the image intensity is greater than that constant.

2.3. Detect plates and characters

After preprocessing, we have the binary image. Now, we begin to recognize characters and plates in binary image. The function retrieves contours from the binary image using the algorithm [1]. The contours are a useful tool for shape analysis and object detection and recognition.

Next, we check all contours to find which contour might be a possible character by checking its width, its height, ratio of its width and its height in a specific range. Finish this step, we have a list of possible character in image.

We arrange all possible character in list by its center, its diagonal length, its width and height. Those possible characters have similar above criteria will be grouped into plates. Then, we have a list of possible plates. Within each plate, we have to sort the characters from left to right by its position.

Finally, each plate is a list of possible character. We apply KNN Algorithms to detect those character, then we have a string represent the text found in possible plate.

2.4. Validate Result

After gotten a list of possible plate with a text inside, we have to validate whether it is a valid license plate. First, we use regular expression to check if the text satisfies the format 00X-0000 or 00X-000.00 (which are formats of license plate in Viet Nam). Then we

have to check the number is exactly one of existing license plate by query database of license plate in Viet Nam. By this validation, we could get the valid license plate in image streaming from camera.

IV. ANALYSIS

Our analysis is partly heuristic – we back it up with experiments reported in the next section.

1. The performance of beacon devices

Beacon devices use Bluetooth 4.0 which has following improvement from classical Bluetooth:
Power Consumption: Bluetooth LE, as the name hints, has low energy requirements. It can last up to 3 years on a single coin cell battery.

Lower Cost: BLE is 60-80% cheaper than traditional Bluetooth.

Application: BLE is ideal for simple applications requiring small periodic transfers of data. Classic Bluetooth is preferred for more complex applications requiring consistent communication and more data throughput.

Apple's iBeacon standard calls for an optimal broadcast interval of 100 ms. Broadcasting more frequently uses more battery life but allows for quicker discovery by smartphones and other listening devices. Standard BLE has a broadcast range of up to 100 meters, which make Beacons able to outdoor location tracking and awareness.

We could set up broadcast range of beacon about 2 meters and install it in the middle of each lane, (lane width is about 3-4 meters). Therefore, a vehicle running with velocity 30-40km/h ($\approx 8.3\text{-}11\text{m/s}$) will pass the region of beacon in 300-500ms. Thus, we will set the broadcast interval of beacon to less than 300ms. We choose 100ms for better signal.

2. The accuracy of license plate recognition

KNN is a special case of a variable-bandwidth, kernel density “balloon” estimator with a uniform kernel.[2][3]

The naive version of the algorithm is easy to implement by computing the distances from the test example to all stored examples, but it is computationally intensive for large training sets. Using an approximate nearest neighbor search algorithm makes k-NN computationally tractable even for large data sets. Many nearest neighbor search algorithms have been proposed over the years; these generally seek to reduce the number of distance evaluations actually performed.

KNN has some strong consistency results. As the amount of data approaches infinity, the two-class k-NN algorithm is guaranteed to yield an error rate no worse than twice the Bayes error rate (the minimum achievable error rate given the distribution

of the data).[4] Various improvements to the k-NN speed are possible by using proximity graphs.[5] For multi-class KNN classification, Cover and Hart (1967) prove an upper bound error rate of

$$R^* < R_{KNN} < R^* \left(2 - \frac{MR^*}{M-1} \right)$$

where R^* is the Bayes error rate (which is the minimal error rate possible), R_{KNN} is the k-NN error rate, and M is the number of classes in the problem. For $M=2$ and as the Bayesian error rate R^* approaches zero, this limit reduces to “not more than twice the Bayesian error rate”.

There are many results on the error rate of the k nearest neighbor classifiers.[6] The k-nearest neighbor classifier is strongly (that is for any joint distribution on (X, Y)) consistent provided $k = k_n$ diverges and k_m/n converges to zero as.

The K-nearest neighbor classification performance can often be significantly improved through (supervised) metric learning. Popular algorithms are neighborhood components analysis and large margin nearest neighbor. Supervised metric learning algorithms use the label information to learn a new metric or pseudo-metric.

V. EXPERIMENTAL RESULTS AND CONCLUSION

After section 3, we build the database and configure beacon packet and broadcast time interval. In experiment, mobile app received packet from beacon within 200ms. When mobile app analyze the packet, it got exactly important three field UUID, Major and Minor with 100% correctness. With each combination of those three field, it request the information of toll station correctly. Therefore, we believe that this way using beacon and BLE is completely able to be applied for tracking the toll station from driver's mobile app.

Using KNN algorithms to recognize the license plate is a good solution here. However, with the limit of samples, the accuracy is not high. There are 68.3% of photos has license plate was recognized correctly. We are choosing K=1 with two sample (include 24 characters A to Z, and 10 numbers 0 to 9).

In future, we could enhance its performance by adding more sample and choose a higher K for this algorithms.

REFERENCES

- [1] Satoshi Suzuki and others. Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing, 30(1):32–46, 1985.
- [2] D. G. Terrell; D. W. Scott (1992). “Variable

- kernel density estimation". *Annals of Statistics*. 20 (3): 1236–1265.
- [3] Mills, Peter. "Efficient statistical classification of satellite measurements". *International Journal of Remote Sensing*.
- [4] Cover TM, Hart PE (1967). "Nearest neighbor pattern classification". *IEEE Transactions on Information Theory*. 13 (1): 21–27.
- [5] Toussaint GT (April 2005). "Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining". *International Journal of Computational Geometry and Applications*. 15 (2): 101–150.
- [6] Devroye, L., Gyorfi, L. & Lugosi, G. (1996). *A probabilistic theory of pattern recognition*. Springer.

Apply Collaborative Filtering to give Suggestions for Users in Web Application

Nguyen Viet Cong Hau

FPT University
HCMC, Viet Nam

Luu Nguyen Ngoc Thu

FPT University
HCMC, Viet Nam

Nguyen Ngo Duc Khai

FPT University
HCMC, Viet Nam

Abstract

Nowadays, websites are one of the most important things in human life. People access and go online every day, every hour, every hour and every second. However, the total information in the “online world” - world of websites is too much. Users have many choices to view on the web. How do users select suitable item to view in dozens, hundreds or thousands (or more) items in a website? How do website owners can keep user staying with them if they do not give “the right item” for “the right people”? That is why websites need to have some ways to give recommended items to their users. These items should be something that users like. We can see the example for this case every day. For example, Amazon suggests products that people may like to buy, or products like or related to the product that a user is browsing. Another example is the suggestions of Facebook’s advertising posts (or sponsor posts) that its user may be comfortable to see. The more items users may like displayed, the more chance to keep they stay the website.

Keywords

Recommendation; suggestion; collaborative filtering; pearson correlation score; recommended item

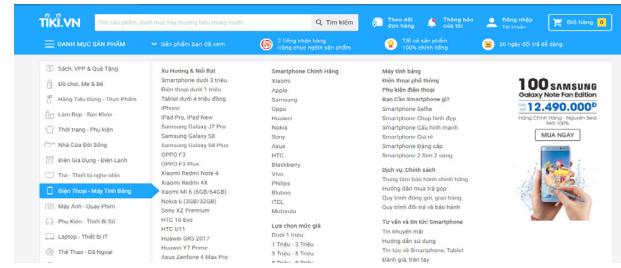
I. INTRODUCTION

The basic method that a website may use to display its items to users is category filtering. For example, Tiki.vn is an ecommerce website. They sell many products. To help users target that they want to buy, products are listed in to category. A category may be a child of another category to grouping related categories.

To look for a product by this method, users need to start from the “root” category, then choose the sub-category if any, and then select a product to browse.

Assume that a user would like to buy an Iphone, he or she needs to do following steps of category filtering method:

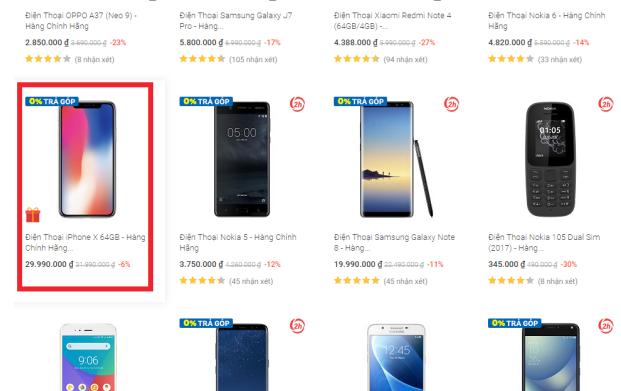
- Step 1: at the Tiki.vn homepage, click the category “Điện thoại – máy tính bảng” (English: Cellphones – Tablets).



- Step 2: After the browser redirect to the category “Điện thoại – máy tính bảng” page, click to the sub-category “Điện thoại smart phone” (English: smart phones).



- Step 3: After the browser redirect to the category “Điện thoại smart phone” page, user is now able to browse the product “Iphone” smart phones.



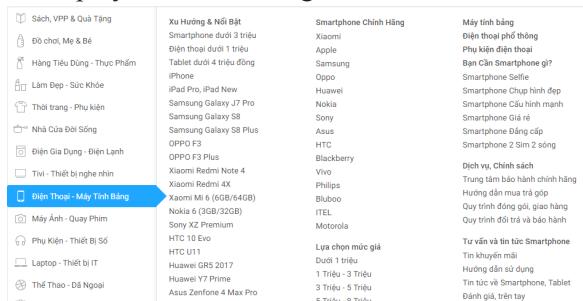
It seems good because now the user can buy a new Iphone. However, user needs to do many steps to get the right product. This method may be good if user really know what they want. It can start from his or

Supervisor: Nguyen Huy Hung, M.E
(Head of SE Department, HCMC FPT University)
Defended topic in front of the scientific council of HCMC FPT University on December 13th, 2017
Accepted post on December 26th, 2017

her willing: Buy a new Iphone. What is an Iphone? It is a “smart phone”. Then, the point is now should list the products by the category “smart phone”. What if user did not really know what the exactly product they would like to buy? What if user just “want to buy something”?

We assume that user would like to buy a “cellphone”. If you were Tiki.vn owner, which product would you like to sell to this customer? A “feature phone” or a smart phone? A “cheap smart phone” or a “flagship” smartphone?

Turn back to the homepage of Tiki.vn, if you hover the mouse (do not click) to the category “Điện thoại – máy tính bảng”, what you get? A pop-up menu will be displayed like the image below:



We could see that the top column is “Xu hướng & Nổi bật” (English: Trends & Highlights).

Xu Hướng & Nổi Bật

[Smartphone dưới 3 triệu](#)

[Điện thoại dưới 1 triệu](#)

[Tablet dưới 4 triệu đồng](#)

[iPhone](#)

[iPad Pro, iPad New](#)

[Samsung Galaxy J7 Pro](#)

[Samsung Galaxy S8](#)

[Samsung Galaxy S8 Plus](#)

[OPPO F3](#)

[OPPO F3 Plus](#)

[Xiaomi Redmi Note 4](#)

[Xiaomi Redmi 4X](#)

[Xiaomi Mi 6 \(6GB/64GB\)](#)

[Nokia 6 \(3GB/32GB\)](#)

[Sony XZ Premium](#)

[HTC 10 Evo](#)

[HTC U11](#)

[Huawei GR5 2017](#)

[Huawei Y7 Prime](#)

[Asus Zenfone 4 Max Pro](#)

Tiki.vn is suggesting us to buy “Smartphone dưới 3 triệu” (English: Smartphones which’s price less than 3.000.000 vnd), “Điện thoại dưới 1 triệu” (English: Cellphones which’s price less than 1.000.000 vnd), “Tablet dưới 4 triệu đồng” (English: “Tablets

which’s price less than 4.000.000 vnd), iPhone, iPad Pro, iPad New, etc.

It seems that there are product categories that has most searching times from the other users, or just simply categories which’s products is “bestselling” products.

Is recommending “bestselling” products or top search products an effective way to make customer buy more?

II. PROBLEM AND SOLUTION PLAN

By using category filtering method or trending listing may not an effective way to make user stay more at the website.

The category method need user target what they really want and take time to browse items. On the other hand, the trending listing take less time, but the trending suggestions may be not the items user like to browse.

We would like to introduce another effective method to give suitable recommended items to website’s users. It is a method introduced by Toby Segaran in the book “Programming Collective Intelligence” published on August 2007 by O’Reilly to solve this problem. It is the method Collaborative Filtering.

It is the method that give recommended items to a user base on the data from the other users who are similar to the user.

To understand this method, we need to think about the case of recommending in real life. Assume that the user in the example of section I would like to buy a new “smart phone”. If he or she does not find products on Tiki.vn, what could he or she do? He or she may ask some friends to give him or her suggested phones to buy. But, how he or she could ask a friend on a website? The website may not know who is friend of the user, but there is an effective way to give suggested items to user based on the other experience.

That method is finding the other users that have “same taste” with the user. When two people has same taste or similar together, we may use this person’s experience to give recommendations to the other.

III. PLAN IMPLEMENTATION

With Collaborative Filtering, we need to do these following steps to suggest items to a user:

- Step 1: Build a dataset that is the feedback score users give to items.

The best way to collect experience of users about items is let users give score to item. The more score user give to an item, the more they like the item.

- Step 2: Find similar users by looking for users that has similar feedback to a common (between 2 users)

item set.

To determine the similar user to a user, we should calculate the pearson correlation score (similar pearson score) between the 2 (two) users.

Similar Pearson =

$$= \frac{\sum_{i \in I_{xy}} (r_{x,i} * r_{y,i}) - (\sum_{i \in I_{xy}} (r_{x,i}) * \sum_{i \in I_{xy}} (r_{y,i}) / n)}{\sqrt{\left(\sum_{i \in I_{xy}} (r_{x,i})^2 - \frac{(\sum_{i \in I_{xy}} (r_{x,i}))^2}{n} \right) * \left(\sum_{i \in I_{xy}} (r_{y,i})^2 - \frac{(\sum_{i \in I_{xy}} (r_{y,i}))^2}{n} \right)}}$$

Where:

- I_{xy} is the set of common products that the 2 (two) users have given scores.

- r_x the set of common products' score given by user 1.

- r_y the set of common products' score given by user 2.

- n the numbers of common products that 2 (two) users have given scores.

- Step 3: Suggest items for a user by his or her other similar users. These items are the item that the user has not given scores yet.

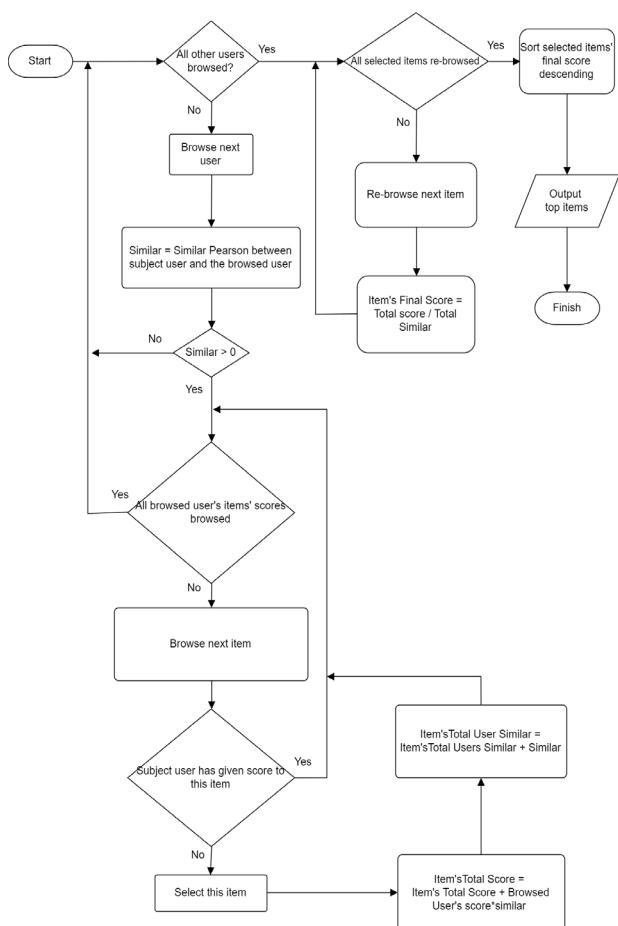
In step 3, we could give out 2 (two) ways to implement.

1. Get the most similar user to the user, give out the recommended list of items that this similar user given highest scores and the user has not given score for this item yet.

But the method 3.1 may not the best way. The reason is that it is too permissive. Therefore, we suggest doing the method 3.2,

2. Get a set of similar users for the user, then calculate the weighted score of items that given score by these users and the user also has not given score for these items. We can list the items that have weighted score larger than 0 as the recommended list of items.

Here is the algorithm for this solution:



IV. ANALYSIS

In total, the complexity of the user-based recommendation is N^2 .

V. EXPERIMENTAL RESULTS AND CONCLUSION

To understand more this method, we should use the example below.

The target is now give suggested products by the collaborative filtering.

Assume that we have a dataset of users' feedbacks' scores to products each time they buy a product as the table below:

User	Product	Score
1	1	0.844444
1	2	3.8
1	3	3.15
1	5	4.35
1	8	3.433333
2	1	2.833333
2	2	4
2	3	3.35
2	9	3.9
3	1	2.9
3	2	4.65

3	3	3.433333
3	6	3.4
3	9	3.65
4	3	3.266667
4	5	4.575
4	8	3.6
4	9	3.8
5	11	3.7
5	19	4.3
5	9	1.2
6	3	4.7
6	12	1.2
6	44	4.5
7	43	2.8
7	44	3.9

Assume that the user needs to be suggested items is user 4. Then we use the scores above to calculate the Pearson correlation score. Then, we could get the similar users to the user 4 as the list below.

User	Similar with User 4
1	0.999803
2	1
3	1

Finally, we list out the scores of all other users give to products that user 4 has not given scores yet.

Expert

User	1	2	3	5
1	0.8444	3.8	3.15	4.35
2	2.8333	4	3.35	
3	2.9	4.65	3.4333	
Total	6.5777 7778	12.45		
Sim Sum	2.9998 0321	2.999 803213		
Total/Sim Sum	2.1927 3643	4.150 27224		

In the above result, we could list out the suggested products for user 4 are: product 2, product 1.

REFERENCES

Programming Collective Intelligence

By Toby Segaran, published on August 2007 by O'Reilly. Copyright © 2007 Toby Segaran. All rights reserved

ISBN-10: 0-596-52932.

ISBN-13: 978-0-596-52932-1

Building Crowd Monitoring of Expo System using iBeacon Technology

Ngo Quang Tuan Anh

FPT University

HCMC, Viet Nam

Vo Thanh Danh

FPT University

HCMC, Viet Nam

Nguyen Ngo Duc Khai

FPT University

HCMC, Viet Nam

Abstract

When an exposition is held, the exhibition organizers often want to know visitors' trends at exhibition to make the strategic decisions at any time. To solve this problem, we introduce a system that provide, collect information about visitors' trends at exhibition. The system includes numbers of beacons, some mobile devices and a web application. We use beacon combined with mobile application to determine the position of the visitors in the exhibition. We collect all user location information to generate the real-time heat map and reports about the visitor's trends.

Keywords

iBeacon; Bluetooth Low Energy; Expo; Proximity Estimote Beacon; Heat map

I. INTRODUCTION

When an exposition is held, there are a number of stalls with a lot of visitors and some with few visitors. The exhibition organizers often count the attractiveness of each stall by eye at a specific time that the organizer decides. This statistic is made subjective at a specific time that the organizer decides. This makes the statistics and analysis of user trends in the exhibition is inaccurate. Beside that, they can be made by survey with visitors, or counting the number of given gifts. However, the exhibition organizers can't force all visitors do survey and control the delivery gifts equally.

In the process of analysis, we believe the beacons technology is capable to complete goals that we set out with the ability to build a device grid that can indicate visitors' trends at exhibition.

The system includes the numbers of beacons, some mobile devices and a web application.

We use beacon combined with mobile application to collect the number of the visitors in the exhibition at any time. From these figures, we can calculate and analysis to generate the heat map and report for expo organizer.

II. PROBLEM AND SOLUTION PLAN

1. How to collect the number of visitors for each stall?

According to the technology researches, we found out that Estimote Proximity Beacon [1] is very capable of resolve the current situations about collecting the number of visitors for each stall. We can use a feature of Estimote Proximity Beacon to solve the problem. The basic idea is using the Proximity Beacons to broadcast Bluetooth 4.0 signal in the region of the stall. When the mobile device sees the signal from the beacons, it will send the beacon's information to server. The beacon's information includes: UUID, major and minor [2]. The system will base on beacon's information to define the stall and update the number of visitors for this stall. From the collected figures, the system will generate the real-time heat-map about visitors' trends at exhibition for both expo organizer and visitor. Besides that, the expo organizer can get reports about visitors' trends to make the strategic decisions.

2. How to mapping beacon for a stall in expo map?

Each expo has its expo map and we found out that we can use this expo map to map the beacon to the corresponding stall. By touching on any position in the expo map. We can get the longitude and latitude of this position in expo map. However, in each device, the size of the expo map is very different. So, when we apply those longitude and latitude in the same expo map in different device, the position of stalls changes.

In order to set the stall position correctly in the same expo map in any devices, we will use those two above figures to calculate and get a couple of ratio. How to calculate this couple of ratio will be introduce in the following section.

III. PLAN IMPLEMENTATION

1. Mapping beacon for a stall in expo map

Supervisor: Kieu Trong Khanh, M.E
 (Head of ITS Department, HCMC FPT University)
 Defended topic in front of the scientific council of HCMC
 FPT University on December 13rd, 2017
 Accepted post on December 26th, 2017

To deploy the system in an exhibition, the organizer will initiate the exhibition by providing a map of the exhibition. Then, the staff will use the mobile to mapping beacons with corresponding stalls. The staff will touch the position of the stall that they want to configure in expo map, the mobile application will display a dialog for staff to enter the stall name. After that, the staff will give a beacon next to the mobile. The mobile application will display the nearest beacon that it sees. The mobile application also notifies the staff when this beacon has been used. When the staff select “Mapping”, the mobile application will send the request includes the stall info, stall location and the beacon info to server. The system will response the result to mobile app.

The position of a stall when the staff touches a position in the expo map will includes two figures: locationX and locationY. LocationX is the ratio between latitude of the stall in expo map and width of expo map in this device. LocationY is the ratio between longitude of the stall in expo map and height of expo map in this device.

Based on this two figures, when the web or mobile application load the expo map, they base on the width and the height of expo map in this device to generate the location for all stall in expo map.

2. Collect the number of visitors for each stall

Visitors must use the internet-connected smart phones that support Bluetooth v4.0 (BLE) along with our application installed. Estimote Proximity Beacon provides API [3] for android and iOS smart phones. According to this API, we could set up handler for event “enter” or “exit” a beacon region. In iBeacon technology, beacon region can be defined in three information UUID, major and minor. When device enters beacon region, it receives the radio signal from that beacon. Then, it will send the beacon and device information to server. The system will base on beacon’s information to define the stall and update the number of visitors for this stall.

When visitors enter another beacon region, the beacon and device information also will be sent to server. The serve will base on beacon’s information to define the stall to increase the number of visitors for this stall and decrease the number of visitors for the above stall.

From the collected figures, the system will generate the real-time heat-map about visitors’ trends at exhibition for both expo organizer and visitor. The expo organizer use web application to view heat map. The stall uses mobile to view heat map. Besides that, the expo organizer can get reports about visitors’ trends by web application to make the strategic decisions.

IV. ANALYSIS

We set out building the system with the following objectives.

Objective 1: The ability to collect the number of visitors in each stall at any time.

Objective 2: The ability to map a beacon with a stall in the expo

Objective 3: The ability to get heat-map and report about the user’s trend.

In the practical system, objective 1, 2 and 3 is met

V. EXPERIMENTAL RESULT AND CONCLUSION

After hundreds of times testing to improve the system and eliminate errors, the whole process of The Crowd Monitoring of Expo System now can run smoothly and stably.

We have run the system in the expo in FPT University in Dec 1st 2017. The visitor (student) installed our application and all data about stalls is collected and send to expo organizer.

ACKNOWLEDGMENT

We would like to thank to Mr. Kieu Trong Khanh for helping us defining project scope and improving the functionality.

REFERENCES

- [1] <http://developer.estimote.com/proximity/>
- [2] <http://developer.estimote.com/ibeacon/>
- [3] <https://github.com/Estimote/SDK>

Customer Relationship Management System for QTSC

Nguyen Thanh Tung

FPT University

HCMC, Viet Nam

Huynh Huu Nhan

FPT University

HCMC, Viet Nam

Nguyen Thi Phuong Nhi

FPT University

HCMC, Viet Nam

Tran Nguyen Bao Duy

FPT University

HCMC, Viet Nam

Abstract

Customer relationship management is a term that refers to practices, strategies, and techniques that help QTSC in particular and others company in general to manage and analyze customer interactions and data throughout the customer lifecycle with the goal of improving business relationships with customer, assisting in customer retention and driving sales growth. CRM systems are designed to compile information on customers across different channels -- or points of contact between the customer and the company -- which could include the company's website, telephone, live chat, direct mail, marketing materials and direct meeting. CRM systems can also give customer-facing staff detailed information on customers' personal information, purchase history, buying preferences and concerns.

CRM software consolidates customer information and documents into a single CRM database so business users can more easily access and manage it. The other main functions of this software include recording various customer interactions (over email, phone calls, social media or other channels, depending on system capabilities), automating various workflow processes such as tasks, calendars and alerts, and giving managers the ability to track performance and productivity based on information logged within the system.

I. INTRODUCTION

At present, QTSC is managing its sales relationship with the customer through a standard method such as by document, verbal communication, and some others direct communication. In addition, each department has their own data and available system which makes communication as well as unifying data difficult.

To resolve the difficulty mentioned, we propose a customer relationship management solution for unifying the data and to make customer relationship tracking channel, is the result of the CRM system. Each employee will use the system with functions corresponding to their role.

Our system is developed based on QTSC procedures, which are much wider and more complicated than regular CRM software available on the market. The system is web application, using QTSC server as database storage.

II. PROBLEM AND SOLUTION PLAN

Customer data and sale activities, as well as customer care, are not managing tightly in several companies nowadays. Every business unit in the company has their own specific customer data. Therefore the connection between departments doesn't exist. The customer's data has just been stored in the simple ways (excel). Some information such as transaction data (date time, customer habit) was not stored. Moreover, there's no tool for managing or tracking sales through email, phone call, etc. Business processes are often delayed because of the interacting process and reporting has been a handwork. The process of tracking customer feedback and corrective actions is not standardized and get a highly effective. As far as the information security is concerned, data sharing between business units as well as customer care faces many obstacles. That's what making the sales and after-sales caring process meet much trouble and bring bad effect as well as customer satisfaction is not high.

Therefore, from now on with CRM company could prevent leaking data and also increase performance of the whole company. Staff has to use this system to store customer data, opportunity, quote and etc. So-called, the manager, as well as staff, can track their work very easy. We also provide tools to export sales quote, analysis report based on QTSC needs. In addition, we do provide a marketing system to manage campaigns. And the highlight of the system is the data sharing between employees and departments, help simplifying the amount of data users have to interact with.

Supervisor: Ngo Dang Ha An, M.E

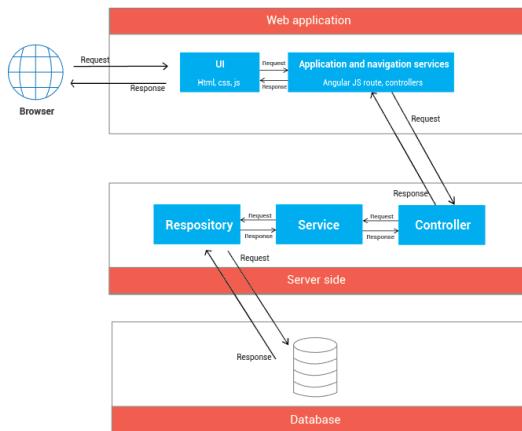
(Head of IA Department, HCMC FPT University)

Defended topic in front of the scientific council of HCMC
FPT University on December 13rd, 2017

Accepted post on December 26th, 2017

III. PLAN IMPLEMENTATION

System architecture:



Description of this architecture:

In Web application, the system is developed under .NET web api architechture style. We choose this architecture for web application becasuse of following advantages:

- Web app contains a web service, with MVC architecture, we can separate business code with Controller and View, so we can use the business code in web service without repeat the code.
 - Web API Controller is the parts of the application that acts like event handler to handler user interaction. Typically, controller read data from a request and calls approriate bussiness's method then selects view to return to user
 - Web service (business logic) is the programming that manages communication between an end user interface and a database. The main components of business logic are business rules and workflows...
- The algorithms are applied to solve the problems:
- Duplicate data checking

1. Definition

Duplicate data checking is an algorithm that checks whether the customer information which customer has given in marketing survey and the customer information in the database are duplicates or not.

2. Defined problem

In most of the marketing campaigns, we have to collect the customer information through a survey. A few years ago, we have to give the survey paper to the customer and ask them to fill in before we add one-by-one survey result to the computer by hand, it really wastes of time. Nowadays, we have google survey, and the customer just needs to fill information on their smartphone or computer and all of these will convey to your system automatically. However, the problem is when they fill information in the electric survey, we cannot control how the survey result would affect to our database. Sometimes,

the customer gives us too much bad information or duplicate data, and we have to spend a huge time to classify that stuff. Therefore, we need something to make this process faster.

3. Solution

To solve this problem, we should follow these steps:

- In this algorithm, we will focus on customer name and their address to calculate. At first, we define a list of string that may appear in customer name such as (TNHH, CÔ Phàn, một thành vién...). As far as we concerned that perhaps these string is consist in the database but not in survey which customer gives and so the system can make a wrong decision in realising two string is similar
- Then we collect all the survey result from customers.
- After that, at each result name, we cut its string into several parts, and remove some strings which are similar to the list we mentioned at the beginning. And we also do the same thing to data in our database.
- Subsequently, we compare each customer given name to our system customer name. And get the results in percent number.
- Next, if the result is 70% matching percent, system return set of records may be matched and continue to compare these records customer address. In this step, we use Google API map distance matrix to find the distance between that set of addresses and the record in the survey. Next, if the result is less than 50m, we will continue to execute the step below to find a 100% percent match, else we will just return the set of record that maybe match
- In this step, we cut the address string of set records into several parts and get only numbers of them. And again we compare that set number like step 4 to get a percent number of similarity. In short, if we will get two similar results. System will return that two records are absolutely the same

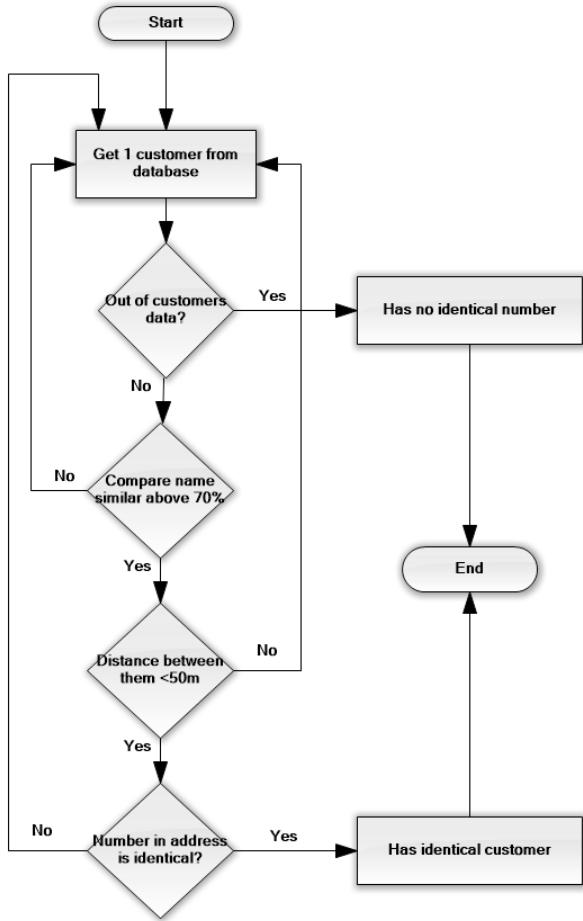
4. Example

Calculate whether two customer is similar:

- Given the customer info in database has name "NINA" and address Lầu 3, Tòa nhà SaigonTel, Số 46, Công viên phần mềm Quang Trung, Phường Tân Chánh Hiệp, Quận 12, TP Hồ Chí Minh
- Give the customer info that customer give in survey has name "NINA" and address "Lô 46, Tòa nhà SaigonTel, Lầu 3, Công viên phần mềm Quang Trung, Phường Tân Chánh Hiệp, Quận 12, TP Hồ Chí Minh"
- Compare two records:
 - NINA ? NINA : 100% matching
 - Distance of address 1 with distance of address 2: 10m
 - Collect number of address 1 and 2, we have: 3,

46, 12 and 46, 3, 12 => 100% matching

- System shows the result that two record is identical
- Flowchart:



IV. ANALYSIS

In the previous sections, we described system architectural design and algorithm. In this section, we are going to introduce why we use this architecture to build our system.

Web API is the great framework for exposing your data and service to different-different devices. Moreover, Web API is open source an ideal platform for building REST-ful services over the .NET Framework. Unlike WCF Rest service, it uses the full features of HTTP (like URIs, request/response headers, caching, versioning, various content formats) and you don't need to define any extra config settings for different devices, unlike WCF Rest service.

Finally, the question is why to choose web API and the answer is below:

1. If we need a Web Service and don't need SOAP, then ASP.Net Web API is the best choice.
2. It is used to build simple, non-SOAP-based HTTP Services on top of existing WCF message pipeline.
3. It doesn't have a tedious and extensive configuration like WCF REST service.

4. Simple service creation with Web API. With WCF REST Services, service creation is difficult.
5. It is only based on HTTP and easy to define, expose and consume in a REST-ful way.
6. It is lightweight architecture and good for devices which have limited bandwidth like smartphones.
7. It is open source.

V. EXPERIMENTAL RESULTS AND CONCLUSION

There are features be tested:

Sales: Login, logout, create customer, create activity, create opportunity, create sale items, change opportunity stage. (PASS: 9/9 test cases)

Support: Login, logout, create issue. (PASS: 4/4 test cases)

Maintenance man: Login, logout, update solving result. (PASS: 4/4 test cases)

Marketing: Login, logout, create campaign, update campaign result. (PASS: 11/12 test cases)

Sales manager: Login, logout, accept quote, view dashboard. (PASS: 5/5 test cases)

Admin: Create account. (PASS: 5/5 test cases)

ACKNOWLEDGMENT

We thank to Mr. Ngo Dang Ha An for helping with the software. We thank the referees of this and the earlier conference version for several suggestions that helped improve the presentation significantly.

REFERENCES

- [1] <http://www.dotnettricks.com/learn/webapi/what-is-web-api-and-why-to-use-it>
- [2] <https://developers.google.com/maps/documentation/distance-matrix/?hl=vi>

DDoS Attack Prevention System

Ma Lam Ngoc
FPT University
HCMC, Viet Nam

To Thanh Thuyen
FPT University
HCMC, Viet Nam

Nguyen Hoang Tuan
FPT University
HCMC, Viet Nam

Abstract

DDoS, or Distributed Denial of Service – is one of the most popular methods of cyber attacks. Despite being popular and recognizable in many media, it is dangerous and hard to control or prevent effectively.

In order to prevent the DDoS attack, many companies use an expensive hardware firewalls, but of course it would cost a lot of money which will cause many problems for the small and medium companies to have their own hardware firewall.

In this paper, we propose a cheaper and more simple method to anti-DDoS attack is to use a build-in firewall in router.

Keywords

pfSense, DDoS, DoS, intrusion prevention system, firewall, anti-DDoS, FreeBSD, Snort, barnyard2.

I. INTRODUCTION

Our team – No Internet Access – has proposed an idea, is to utilize an open source NIDS/IPS (network intrusion detection/prevention system) to specifically deal with DDoS Attacks, with an external service that can learn the traffic pattern via packet then provide countermeasures. It is cost-effective, requires much less effort and much more easy to track progress.

Our team idea is to use the combination of pfSense and Snort, with our own service to manage them.

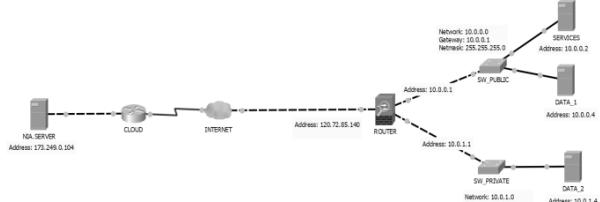


Figure 1. NIA Protection System

Supervisor: Ngo Dang Ha An, M.E
(Head of IA Department, HCMC FPT University)
Defended topic in front of the scientific council of HCMC
FPT University on December 15th, 2017
Accepted post on December 26th, 2017

II. PROBLEM AND SOLUTION PLAN

A Denial-of-Service (DoS) attack is a malicious attempt to render a service, system, or network unusable by its legitimate users. Unlike most other hacks, a Denial of Service (DoS) does not require the attacker to gain access or entry into the targeted server. The primary goal of a DoS attack is instead to deny legitimate users access to the service provided by that server. Attackers achieve their DoS objective by flooding the target until it crashes, becomes unreachable from the outside network, or can no longer handle legitimate traffic. The actual volume of the attack traffic involved depends on the type of attack traffic payload used. With crafted payload such as malformed IP address fragments, several such packets might be sufficient to crash a vulnerable TCP/IP stack; on the other hand, it might take a large volume of perfectly conforming IP address fragments to overwhelm the defragmentation processing in the same TCP/IP stack. Sophisticated attackers might choose to use a mixture of normal and malformed payloads for a DoS attack. DoS attacks can vary in impact from consuming the bandwidth of an entire network, to preventing service use of a single targeted host, or crashing of a single service on the target host.

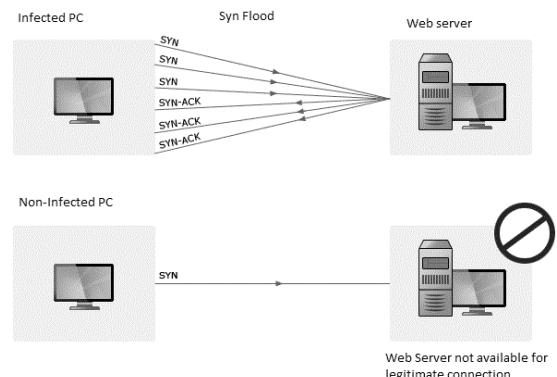
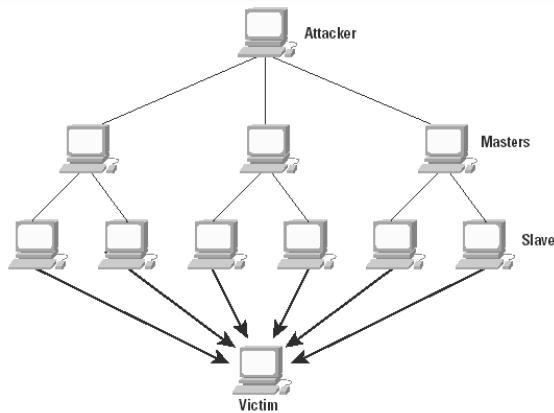


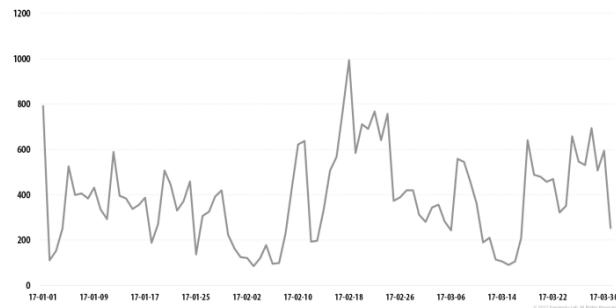
Figure 2. DoS Attack

DDoS, or Distributed Denial of Service – is one of the most popular methods of cyber attacks. Despite being popular and recognizable in many media, it is dangerous and hard to control or prevent effectively.

**Figure 3. DDoS Attack**

According to Kaspersky Lab in Q1 2017 DDoS attack:

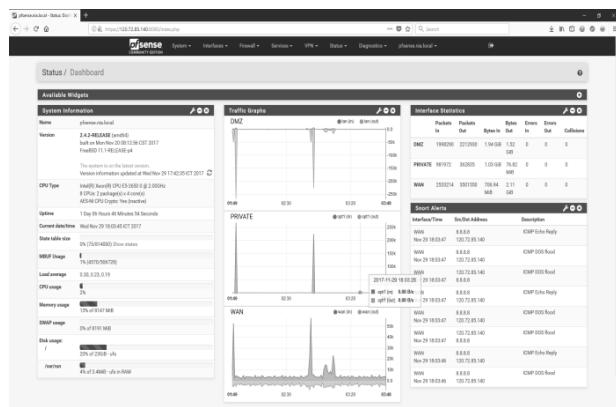
- Resources in 72 countries (vs. 80 in Q4 2016) were targeted by DDoS attacks in Q1 2017.
- 47.78% of targeted resources were located in China which is significantly lower than the previous quarter (71.60%).
- China, South Korea and the US remained leaders in terms of both number of DDoS attacks and number of targets, while the Netherlands replaced China in terms of number of detected servers.
- The longest DDoS attack in Q1 2017 lasted for 120 hours – 59% shorter than the previous quarter's maximum (292 hours). A total of 99.8% of attacks lasted less than 50 hours.
- The proportion of attacks using TCP, UDP and ICMP grew considerably, while the share of SYN DDoS declined from 75.3% in Q4 2016 to 48% in the first quarter of 2017.
- For the first time in a year, activity by Windows-based botnets has exceeded that of Linux botnets, with their share increasing from 25% last quarter to 59.8% in Q1 2017.
- In Q1 2017, the number of attacks per day ranged from 86 to 994. Most attacks occurred on 1 January (793 attacks), 18 February (994) and 20 February (771). The quietest days of Q1 were 3 February (86 attacks), 6 February (95), 7 February (96) and 15 March (91). The overall decline in the number of attacks from the end of January to mid-February, as well as the downturn in March, can be attributed to the decrease in activity by the Xor.DDoS bot family, which made a significant contribution to the statistics.

**Figure 4. Number of DDoS attacks over time in Q1 2017**

III. PLAN IMPLEMENTATION

In order to solve the stated problems, we would suggest to use a router OS called: pfSense.

pfSense software includes a web interface for the configuration of all included components. There is no need for any UNIX knowledge, no need to use the command line for anything, and no need to ever manually edit any rule sets. Users familiar with commercial firewalls catch on to the web interface quickly, though there can be a learning curve for users not familiar with commercial-grade firewalls.

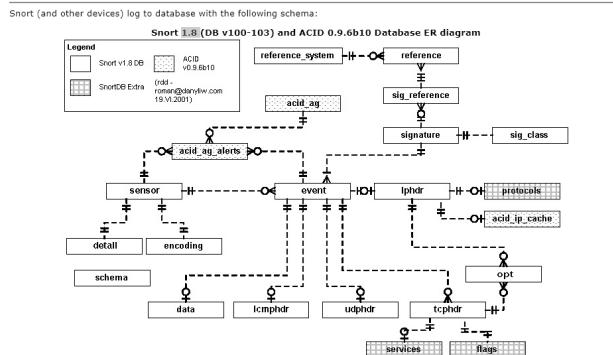
**Figure 5. pfSense Web GUI**

Snort is an open source intrusion prevention and distribution (IPS / IDS) based on Sourcefire's intrusion network. Combined with test signals, protocols and cues, Snort has been deployed worldwide. With monthly downloads and over 400,000 subscribers, Snort has become the standard for intrusion prevention and intrusion detection. Snort's main functions are packet spoofing, packet logging, and network intrusion detection.

Snort is an intrusion detection and prevention system, which can be configured to simply log detected network events to both log and block them. Thanks to OpenAppID detectors and rules, Snort package enables application detection and filtering. The package is available to install from System > Package Manager. Snort operates using detection signatures called rules. Snort rules can be custom

created by the user, or any of several pre-packaged rule sets can be enabled and downloaded. Barnyard2 is a way to store and process the binary outputs from Snort into a MySQL database.

ACID: Database (v100-103) ER Diagram

**Figure 6. Snort Database ER Diagram**

With pfSense & Snort as a bundle, the system is pretty much set. But it's not what our project is entirely about. This is a soft firewall-based protection system, like many others, have flaws in it. So, in an effort to harden the protection and mitigate even more risks, we introduce you to our NIA Service.

Table 1. Plan Implementation

No	Date	Activity
1	9 days	Basic research & laying the foundation to build the test system
2	33 days	Build the system, initialization and basic setup, make sure that the system run smoothly
3	15 days	Perform basic DDoS Attack and fix all problem occurred
4	14 days	Testing the capabilities of the NIA system, perform tests with other types of DDoS
5	3 days	Update the security policy
6	14 days	Finalize the installment, review and demonstration
7	1 day	Train the staff about the system and policy

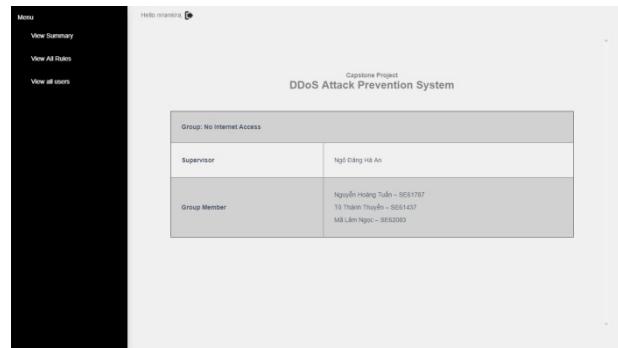
*** Note that all of the activities above can be shorten during implementation progress.

IV. ANALYSIS

The pfSense's firewall has three ways to deal with incoming message: block, reject or pass. With reject, a TCP RST or ICMP port unreachable for UDP is returned to the sender. With block, the packet is dropped silently. If an IP is on the block or reject list, it means that the IP is either consider a member of a botnet or it is one of the botnet.

Our algorithm here is to add all IP that try to reach

out/scan our unsupported ports and services into a library of suspected IPs and those IP will be blocked/rejected when a DDoS attack is happening, and for the IPs only attack one service (HTTP, FTP, etc), we will based on the IPs that connect with the service before/during the DDoS attack to find out who is the proper customer and the rest go to reject/block list. We have two options to deal with the reject/block list, one is completely block them to connect to our service, other is limit their network traffic bandwidth for a period of time. For a small DDoS attack, we will try to find out who is the member of the botnet and limit them. For a larger attack, the only option is to limit all the incoming connection bandwidth. This will make sure that our service and our customer services still be able to run properly.

**Figure 7. Our cloud service main page**

V. EXPERIMENTAL RESULTS AND CONCLUSION

Our solution is consist of two parts: pfSense and NIA cloud services.

A. pfSense

Table 2. pfsense Test Cases And Results

TEST	Meanings - Result
Install pfSense in a virtual machine	Deploy testing environment - Passed
Browse the GUI of pfSense on browser and configure IP and services	Learn how to use the GUI and features its provide and setting the pfSense network interfaces - Passed
Add rules to block, reject, pass for each individual income IP and port, limit income and outcome bandwidth	Testing the capability of firewall in pfsense - Passed
Install Snort and configure it to log all traffic to external mysql database.	These logs will be progressed by our management services - Passed

Deploy pfSense on a cloud machine	Running pfSense as a router on a cloud environment - Passed
Configure DHCP Server	Providing the IP for DMZ and private network - Passed
Deploy business network model with DMZ and private zone.	Experiments on system services - Passed
Browse GUI, testing feature (rules, services, snort, log traffic to mysql)	Make sure pfSense can run on the cloud as same as the virtual machine environment - Passed
Move forward to our management services	

B. Cloud services

Table 3. Cloud services test cases and results

TEST	Meanings - Result
Install JDK 8 Oracle and Tomcat 7 on our Ubuntu 14.04 server	Setting environment for the services server - Passed
Create a web project, build and run for the first time	Make sure the environment has been set up properly - Passed
Create connection to external mysql database	Connection to the database of traffic packages - Passed
Code the login page	Login page for users - Passed
Code the View all users page (admin users only) with CRUD	User management page for admin - Passed
Code the View summary and add rules to pfSense firewall page	Page where user can see summary of the pfSense traffic in the last 24 hours and in this page user can add rule manually to firewall of pfSense - passed
Code the View all rules and edit/delete rules page	Page where user can see all of their existing rules and can decide to delete/edit them.
Test all function again before deploy it to the world	Make sure all function run properly - Passed

ACKNOWLEDGMENT

We would like to take this opportunity to express our profound gratitude towards all the people who have contributed to this project: Our teachers for

their advices and participation in the final review, our friends for their valuable technical support. We would also like to express our appreciation to Mr. Ngo Dang Ha An, our research supervisor for his professional guidance and the useful, constructive recommendations throughout the course of this project.

Secondly we would like to thank our Families and Friends, for their unconditional support and tolerance during the time of doing the project.

REFERENCES

- [1] pfSense Documentation (https://doc.pfsense.org/index.php/Main_Page)
- [2] Snort Official Documentation (<https://www.snort.org/documents>)
- [3] AO Kaspersky Lab [RU] (<https://securelist.com/>)

Dendrobium Orchid Care Solution:

*Deploying in small and medium system to support people
who like taking care of orchids at their home without high cost*

Nguyen Quang Huy
FPT University
HCMC, Viet Nam

Nguyen Tuan Sang
FPT University
HCMC, Viet Nam

Mai Thanh Hieu
FPT University
HCMC, Viet Nam

Huynh Quang Dat
FPT University
HCMC, Viet Nam

Tran Hong Son
FPT University
HCMC, Viet Nam

Abstract

The most garden dendrobium is operated by the farmer or people who like taking care of Dendrobium. In this way, they follow their experience to take care the dendrobium orchid. The traditional bases on human experience. They do not have any machines to measure exactly parameters of environment like humidity, temperature and intensity of light. There are many other third parties have implemented a device to fit our system requirements, but the price is high and all of the vendors are far away from Viet Nam. Some of them do not support or even sell their product outside their country. Therefore, we have decided to research some embedded system and come up with do-it-ourselves solution. There are some Viet Nam companies have provided IoT agriculture devices such as : Bach Khoa, and HACHI. They also provide IoT system for clean agriculture with the product reached VietGAP or further is Global GAP. The solution detail will be described in this paper.

I. INTRODUCTION

In this document, we introduce a solution for anyone who likes Dendrobium orchid and others orchid. We bring the technology to their farm and help them reducing the cost.

We will build a solution. In this solution, we provide and suggest the hardware and the software completely user manual in English and Vietnamese. We provide sensors and controllers to get the temperature, intensity of light and humidity in the garden. We build a mobile application which has functional like tracking temperature and humidity in the garden and controlling water provider.

This document also describes our working process

in 4 months includes our perspective in the system, component designs and detailed core workflows. We hope the system and our solution will help reduce time to take care the orchid. Also we help who is busy, who is newbie and who is love the orchid but they have inexperienced.

II. PROBLEM AND SOLUTION PLAN

Growing orchids is not a science but a cultivated art. Orchid growers are often called artisans, not engineers or technicians. So, this person spread flowers successful, but also the same farmers have no similar results. Why? Simply because each grower's conditions are different. Do you regularly pump water for plants? How do you grow your orchid? Do you plant them in plastic pots where has appropriate environmental conditions. We have researched all documents about current modern garden, analyzed their strengths and weaknesses. We have found out that most of these systems used for large garden with high cost. What is the problem? The people who takes care of their small gardens which has some pots of orchids, do not apply those systems which are not suitable for small garden with high cost. Then, we find factors and give out solutions to solve these problem. This solution is to provide a system to tracking the garden. This is applied IOT which can control from anywhere have the Internet. The mobile application also provides the recommendation for users. How to use fertilizer suitable with the weather and the stage in time life of orchid. This solution uses industrial devices with high quality and reliability. We provide a HDMI monitor to control at the garden and an application for android mobile.

PROBLEM:

As mentioned in the introduction, we are researching and developing the system which is focusing on small model with low cost. This model is suitable for growing some pots of orchids at home which has small garden and balcony. Although they are Dendrobium lovers, but they don't have knowledge, devices and items for taking care of their flowers. Besides these devices are very high cost because they

Supervisor: Nguyen Duc Loi, M.E
(Lecturer of IT Department, HCMC FPT University)
Defended topic in front of the scientific council of HCMC
FPT University on December 14th, 2017
Accepted post on December 26th, 2017

are used for agriculture or larger garden outhouse for production. On the other hands, they do not have any time for growing orchids because they spend time for their work. In this time, environmental conditions are getting worse for their Dendrobium. What should they do in this case?

Some problems for orchid growers:

- **How do I water for orchids?**

This question seems to be a question that does not have the correct answer because each person's cultivation is different, each orchid also adapts to different soil to grow and develop so we can only base on some signs of orchid illnesses to correct the proper irrigation. Many orchids died due to waterlogging because of its owner watering too much. Growing in terracotta pots requires more water than plastic pots. If grown with sunflower and wood, it is necessary to water more water when grown in pots.

- **How do I grow orchids for flowering?**

Many people grow orchids which don't flower as expected. If there are flowers, then it is not right time. They don't know the causes and they forget core problem of growing orchids due to the weather decision to affect orchid flowers in right time. Surely one thing if good orchid care is still flowering, but flowering depending on the different times we can work by increasing the light or appropriate fertilizer for each orchid.

Many of people living in the high-rise residential complex in the big cities in the country such as Hanoi, Ho Chi Minh City certainly encounter these problems. We have received such information "I grow orchids on the 26th floor, you see if I grow orchids which can grow there?, is there no flowers ?" Obviously with this question we can not answer you correctly. According to the logic of all plants and animals live only in the appropriate **conditions to develop well**.

SOLUTION PLAN:

- **What is model suitable for this system?**

Model includes :android application, 7 inch screen, sensors, microcontrollers, server, orchid hanging frame and some other hardware.

We use :

- SHT10 sensor for getting temperature and humidity
- TSL2561 sensor for getting intensity of light
- ESP8266 module 12 Wi-Fi for developing send/receive data through Wi-Fi
- Raspberry Pi 3 with raspbian OS to connect HDMI 7inch screen
- MQTT Server is used for sending / receiving to microcontrollers and android application.
- Orchid handling frame :

1.5m X 0.5m X 1.5m dimension

- Led bulb 3W BN-L503 : x3 bulbs
- Nebulizer for watering
- Stepper motor for pulling cover
- Total cost : about 7 million VND

- **How do growers take care of orchids when they are not at home?**

All parameters of temperature, humidity and intensity of light will be sent from microcontrollers to MQTT server though Wi-Fi. Wherever you are, just an android phone installed our application which can get all data from MQTT server to give some suggestions for status orchids how should be taken care of. You can control led, nebulizer by using this application or control directly with touch screen.

- **Not much knowledge to grow variety of orchids**

We will collect data of different orchid types then give instructions which is to users if they want to grow new orchid type.

- **If intensity of light is so strong, what should we do?**

We get intensity of light by TSL2561, if intensity of light is so strong, it affect orchids worse, we design a grip controlled by stepper motor.

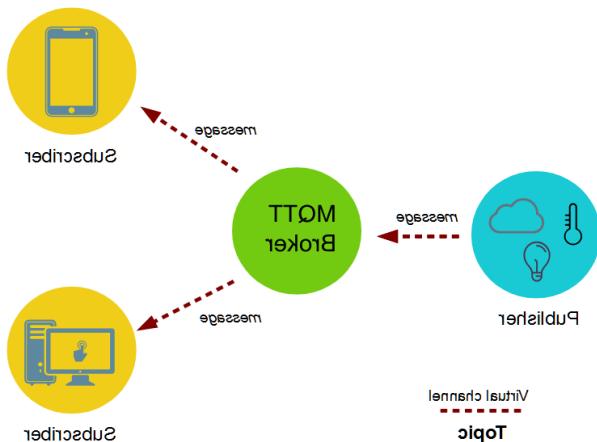
III. PLAN IMPLEMENTATION

1. Control device with central controller

We develop system to have main control board which contain Esp8266 NodeMCU Lua CP2102, Raspberry pi 3... . Esp8266 will read temperature and humidity from STH10 sensor and intensity of light from Lighting sensor TSL2561, then Esp8266 publish data to MQTT Server. Data will be displayed on hdmi 7 inch screen through a library developed by Node Red. Android application will subscribe MQTT Server to get data, and suggest solutions for user.

2. Send and receive with MQTT protocol

MQTT stands for MQ Telemetry Transport. It is a publish/subscribe, extremely simple and lightweight messaging protocol, designed for constrained devices and low-bandwidth, high-latency or unreliable networks. The design principles are to minimise network bandwidth and device resource requirements whilst also attempting to ensure reliability and some degree of assurance of delivery. These principles also turn out to make the protocol ideal of the emerging "machine-to-machine" (M2M) or "Internet of Things" world of connected devices, and for mobile applications where bandwidth and battery power are at a premium.

**Figure 1. MQTT protocol**

3. Communicate HDMI 7inch with Raspberry pi 3

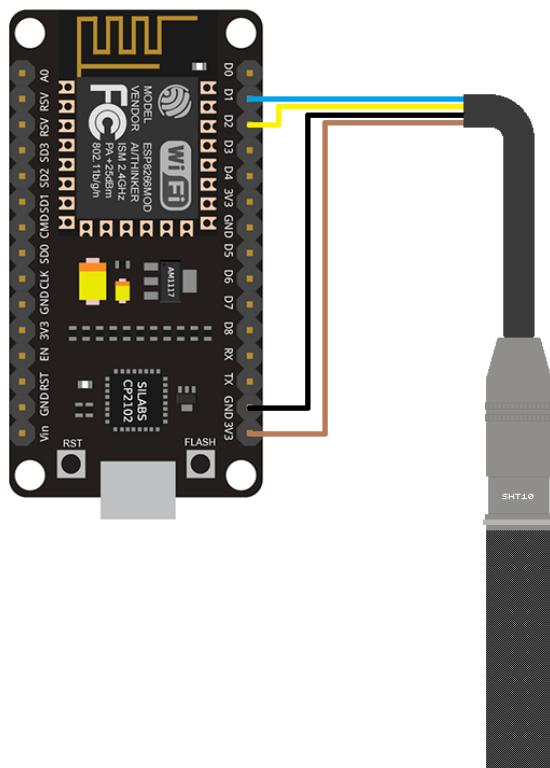
Turn on the “backlight” switch then connect the LCD to your Pi (HDMI Port of LCD => HDMI Port of Pi; USB Port of LCD => USB Port of Pi; 5V~2A power supply). Download the Raspbian image from Raspberry Pi web site. Write the image to a TF card and append the following lines to the config.txt file which is located in the root of your TF card :

```
max_usb_current=1
hdmi_group=2
hdmi_mode=87
hdmi_cvt 800 480 60 6 0 0 0
hdmi_drive=1
```

Save and connect the TF card to your Pi then power up.

4. Communicate STH10 sensor with ESP8266

The temperature and humidity sensor has 3 wires : red, black or green, yellow and blue. The red wire connects to VCC (3-5VDC), black or green wire connects to GRD, yellow wire connects to Clock and blue wire to connects to Data pin on ESP8266

**Figure 2. Communicate SHT10 with ESP8266**

5. Communicate ESP8266 with MQTT Server

NodeMCU Amica is a ESP8266 Wi-Fi Module based development board.

PubSubClient is an Arduino-based MQTT client. With just a few lines of code, it is very easy to either subscribe to topics or publish new ones.

In our setup, we are using the Arduino IDE to program ESP8266

6. Develop interface for displaying

data on HDMI 7inch Screen with
Raspberry pi 3 by Node Red

Node-RED is a programming tool for wiring together hardware devices, APIs and online services in new and interesting ways.

It provides a browser-based editor that makes it easy to wire together using the wide range of nodes in the palette that can be deployed to its runtime in a single-click. So it is very suitable for our project, it support to create an simple interface and useable for user. Besides raspberry pi 3 use raspian OS so it is seen as a minimize computer, it is very convenient for accessing web service to show all data for user on hdmi 7 inch screen.

Users can control directly controller by touching buttons on touch 7 inch screen, the raspberry pi 3 will access API which provided by Node Red . An interface to display all digital data (temperature, humidity, intensity) graphically. This interface is based on Node Red, this time it is seen as MQTT

client, it is like android application which get all data from MQTT server.



Figure 3. Three orchid zones's data on touch screen

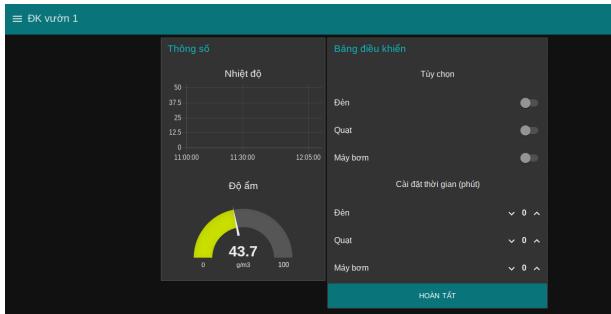


Figure 4. Chart of temperature, humidity and control board

7. Communicate android application with MQTT Server

Android application is seen as MQTT client so if it get data from MQTT Server, it must subscribe the topics of server. Besides it may send data to MQTT by publish method.

We build an android application for user to control and manage stages of dendro. At the first time, the application will connect to MQTT server to get sensor SHT10 's information that were sent through ESP8266. Depend on received data, this application will provide and suggest some solutions for taking care of their dendro such as : how much water , how intensity of light, view history, chart...

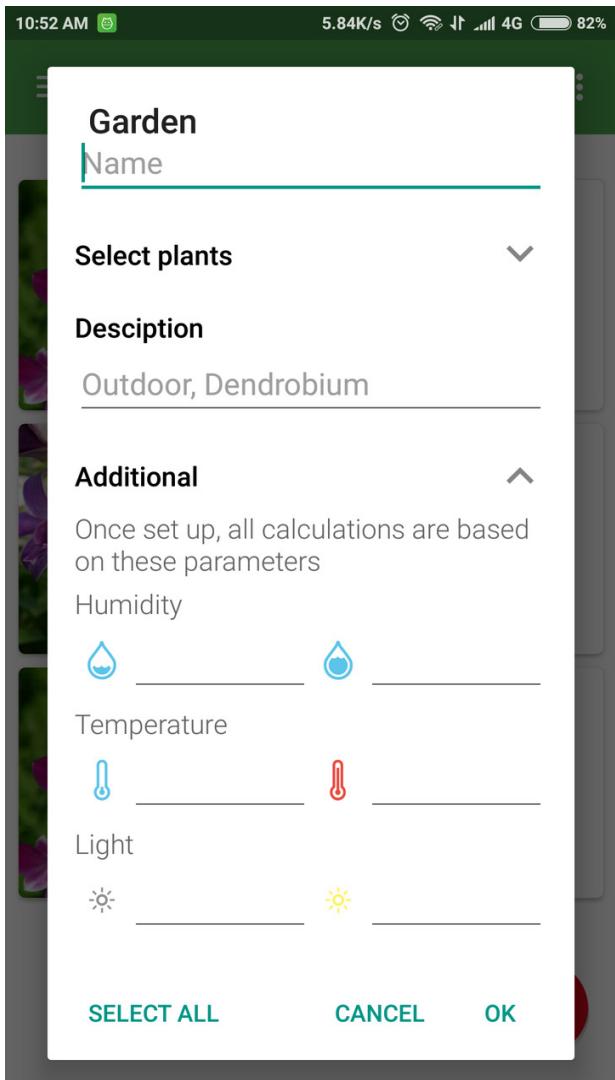


Figure 5. Interface of android app

8. Communicate Lighting sensor TSL2561 with ESP8266

TSL2561 sensor used to measure the intensity of light usually, infrared in the lux unit with high stability and accuracy, the sensor has internal ADC and preprocessor so the value returned is the direct value of light intensity Lux without any processing or calculation via I2C interface. You will connect four of the five pins on the board to ESP8266. The four pins you need are labeled 3V3, GND, SCL and SDA

The fifth pin, INT is an optional interrupt signal which the TSL2561 can use to “interrupt” your microcontroller. You can set up the TSL2561 to automatically send an interrupt when it completes a measurement, or if a measurement goes above or below a certain level for a certain amount of time. This pin is not needed for the basic operation of the TSL2561.

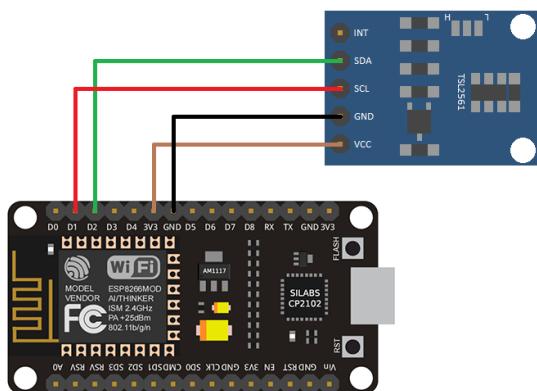


Figure 6. Communicate TSL2561 with ESP8266

IV. ANALYSIS

1. Overview about one zone of system

This system has 3 zones of state different, every zone has 4 orchid pots and 1 led bulb, TSL2561 sensor and 1 SHT10 sensor controlled by ESP8266 through relay module high/low level trigger.

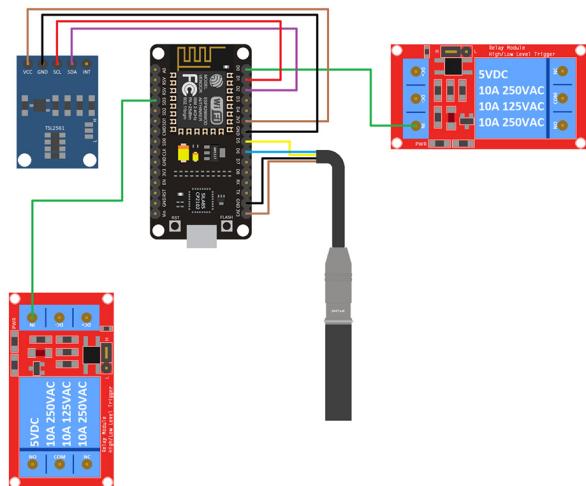


Figure 7. ESP8266 communicate with 2 relay, SHT10 and TSL2561

2. The internet is essential to the system

All mentioned above, our system will revolve around MQTT server, all data will be published and subscribed in the server. ESP8266 will get data from SHT10 and STL2561, then those data are published to topic of MQTT server (depend on the type of data will be included in different topics such as : /topic/temperature, /topic/humidity, /topic/intensity). In our system, raspberry pi 3 and android application must subscribe these topics. After subscribing topics, pi 3 will show data in touch 7 inch screen. So our system will depend on the internet.

Nowadays, most apartments have their own Wi-Fi and companies where they are working too. They can use Internet everywhere. Internet is not important issue for this system.

3. Celsius to Fahrenheit Conversion

In android application, users can view data of temperature by Celsius or Fahrenheit, so the convert function was implemented :

We have 0 degrees Celsius is equal to 32 degrees Fahrenheit ($0^{\circ}\text{C} = 32^{\circ}\text{F}$):

The temperature T in degrees Fahrenheit ($^{\circ}\text{F}$) is equal to the temperature T in degrees Celsius ($^{\circ}\text{C}$) times $9/5$ plus 32:

$$T_{(^{\circ}\text{F})} = T_{(^{\circ}\text{C})} \times 9/5 + 32 \text{ or } T_{(^{\circ}\text{F})} = T_{(^{\circ}\text{C})} \times 1.8 + 32$$

4. ESP8266 Node MCU with CP2102

The CP2102 is a USB to UART chip from Slabs. The CP2102 is compact in size and requires very little external components to operate right away. The module has a 3.3V output voltage with a maximum output current of 100mA.

On the circuit there are 5 output ports: 3.3V RST 5V Tx Rx Gnd. Where RST pins are used to disconnect the module's USB connection to the computer. Maximum speed is 115200 bps, so esp8266 can receives data from sensors quickly and accurately.

4. Voltage for every type of devices and hardware

In our system, we use :

- Led bulb 220V-3W x3
- Esp8266 nodemcu 5V x3
- Relay module high/low level trigger 3.3V x3
- Stepper motor 12V
- Nebulizer 24V

V. FUTURE PLAN

- There are no perfect solutions to problems, as well as there are no perfect systems. With the inexperience of our team members and the time constrains, our proposed solution and project contains many issues. Below are the problems encountered in this project:
 - Hardware Knowledge: We are inexperienced with hardware. All the hardware components chosen to be used in this project is based on our familiar with them, or based on the shortest time we need to learn how to use them. So they are only the most appropriate, not the best choice for the project.
 - Orchid Hanging Frame : It is designed and wired appealingly.
 - Imperfect cost : We have not much time to research many types of sensors, so we haven't chosen device and hardware with perfect cost yet.
 - Android Application : Many functions are not perfect and need some additional functionality in the future.
- Our future plan is try to solve these problems one by one.
- We will also try to develop this system with PLC for large models or agricultural production.

REFERENCES

- [1] <https://www.raspberrypi.org/products/raspberry-pi-3-model-b/>
- [2] <http://hshop.vn/products/kit-rf-thu-phat-wifi-esp8266-nodemcu>
- [3] <https://en.wikipedia.org/wiki/MQTT>
- [4] <https://nodered.org/>
- [5] <https://www.baldengineer.com/mqtt-tutorial.html>

Develop a Security Information Event Management System

Le Xuan Minh
FPT University
HCMC, Viet Nam

Vo Thong Minh
FPT University
HCMC, Viet Nam

Tran Quoc Phuong Nam
FPT University
HCMC, Viet Nam

Do Duy Thinh
FPT University
HCMC, Viet Nam

Abstract

Nowadays, life has become more and more advantages but along with it there are some new issues emerged, one of them is cybercrime. Cyberattack can happen in many forms and in many ways but no matter what method the criminals used, there always have trace left with every system can collect by default is log. But to monitor all kind of log that has been created by system can be a pain since there will have many type of logs be created along with alert logs. To solve that problem effective and cost less as possible, a system, which is a combination of many open source security programs under follow a standard SIEM model, can be used. This security system can help monitoring log easier and allow user to know any abnormal happen in primary system.

Keywords

Security Information and Event Management(SIEM); Security Information Management (SIM); Security Event Management (SEM); Host Intrusion Detection System (HIDS); Network Intrusion Detection System (NIDS)

I. INTRODUCTION

Cybercrime is a fast-growing area of crime. More and more criminals are exploiting the speed, convenience and anonymity of the Internet to commit a diverse range of criminal activities that know no borders, either physical or virtual, cause serious harm and pose very real threats to victims worldwide [1].

A successful attack on computer systems rarely look like real attacks except in hindsight – In fact, if you didn't know better, your own network and system Admins might look a lot like hackers. They are often using elevated privileges to make changes that could look (or be) malicious [2].

Most systems and software generate logs including operating systems, Internet browsers, sale systems,

workstations, anti-malware, firewalls, and intrusion detection systems (IDS) which can be used to know whether system has been attack and trace back to the attacker. Some systems with logging capabilities do not automatically enable logging so it is important to ensure all systems have logs turned on. Some systems generate logs but don't provide event log management solutions [3].

The biggest problem with logs is that nobody cares or look at them. From a security point of view, the purpose of a log is to act as a red flag when something bad is happening. Reviewing logs regularly could help identify malicious attacks on your system. Given the large of amount of log data generated by systems, it is impractical to review all of these logs manually each day. Log monitoring software takes care of that task by using rules to automate the review of these logs and only point out events that might represent problems or threats. Often this is done by using real-time reporting systems that alert you via email or text when something suspicious is detected.

One way to do this is to use SIM (Security Information Management) which is the practice of collecting, monitoring and analyzing security-related data from computer logs or a type of software that automates the collection of event log data from security devices, such as such as firewalls, proxy servers, intrusion-detection systems and antivirus software.[4]

Therefore, SIM needs to combine with SEM (Security Event Management) which has the ability for providing strong event management, real-time threat analysis, visualization, ticketing, incident response, and security operations. They are typically based on enterprise SQL databases such as Oracle. [5]

II. PROBLEM AND SOLUTION PLAN

However, both types have its own disadvantages that may hinder the effectiveness of log monitoring process. For SIM, they are not ideal for complex alerting and not good for security (incident) reporting, trending or dashboards. SIM uses historical analysis and reporting for security event data. This requires event and data collection/correlation (not in real time), an indexed repository for log data and flexible

Supervisor: Ngo Dang Ha An, M.E
(Head of IA Department, HCMC FPT University)
Defended topic in front of the scientific council of HCMC
FPT University on December 15th, 2017
Accepted post on December 26th, 2017

query and reporting capabilities. And for SEM, they are slow when producing reports and rely on a massive index to allow for database queries.

Therefore, we – Group 1 find and combine the open source SIM and SEM software as a SIEM method to make the most secure environment to our business. We don't use open source SIEM simply because most the advanced technology option was locked as only a fee for monthly subscription to use those function which includes log management, cloud based environment, logger, etc... These fees can be from 1000 – 2000 USD/month or 6000 USD/nodes, which most of small to medium business cannot afford.

SIEM technology provides real-time analysis of security alerts generated by network hardware and applications. SIEM solutions come as software, appliances or managed services, and are also used to log security data and generate reports for compliance purposes.

SIEM describes the product capabilities of gathering, analyzing and presenting information from network and security devices; identity and access management applications; vulnerability management and policy compliance tools; operating system, database and application logs; and external threat data.

A key focus is to monitor and help manage user and service privileges, directory services and other system configuration changes; as well as providing log auditing and review and incident response.

III. PLAN IMPLEMENTATION

Our system is a combination of SIEM components which include OSSEC, Snort and Graylog.

1. OSSEC

The OSSEC HIDS is able to play a role in defending the system on which it is hosted. However, as with virtually any intrusion detection software, it can only notify you during or after the attack has already happened. It is vital that an OSSEC HIDS server be one of the best-defended systems in the network. The operating system must help defend the OSSEC HIDS installation so the OSSEC HIDS can in turn help defend the network.

1.1. System Integrity Check (syscheck)

File integrity checking is an essential part of all HIDS technologies and is typically performed by comparing the cryptographic checksum of a known good file against the checksum of the file after it has been modified. The OSSEC HIDS looks for changes in the MD5/SHA1 checksums of designated files on the system and registry keys within the Windows registry. The OSSEC HIDS agent scans the system every few hours, or at an interval you specify, and it sends the checksums of the monitored files and

registry keys to the OSSEC HIDS server. The server stores the checksums and looks for modifications by comparing the newly received checksums against the historical checksum values of that file or registry key. An alert is sent if anything changes [6].

1.2. Registry Monitor

The system registry is a directory listing of all hardware and software settings, operating system configurations, and users, groups, and preferences on a Microsoft Windows system. Changes made by users and administrators to the system are recorded in the system registry keys so that the changes are saved when the user logs out or the system is rebooted. The registry also allows you to look at how the system kernel interacts with hardware and software. An HIDS can watch for these changes to important registry keys to ensure that a user or application isn't installing a new or modifying an existing program with malicious intent [7].

1.3. Rootkit Detection

In addition to the log analysis and integrity checking capabilities previously discussed, the OSSEC HIDS rootcheck process performs rootkit detection on Linux, Unix, and BSD systems. Rootkit detection on Linux, Unix, and BSD systems finds application-level rootkits through signature definitions and kernel-level rootkits using system call comparisons. Anomaly based checks are also performed to ensure that the system is operating as anticipated.

1.4. Enforcing/Monitoring Policies

OSSEC also performs policy monitoring/enforcement on Windows, Linux, Unix, and BSD systems. Policy monitoring/enforcement is the process of verifying that all systems conform to a set of pre-defined policies surrounding configuration settings and approved application usage.

2. Snort

Snort is an open source Network Intrusion Detection System (NIDS) which is available free of cost. NIDS is the type of Intrusion Detection System (IDS) that is used for scanning data flowing on the network.

2.1. Snort Detect Engine

The detection engine is the most important part of Snort. Its responsibility is to detect if any intrusion activity exists in a packet. The detection engine employs Snort rules for this purpose. The rules are read into internal data structures or chains where they are matched against all packets. If a packet matches any rule, appropriate action is taken; otherwise the packet is dropped. Appropriate actions may be logging the packet or generating alerts. The detection engine is the time-critical part of Snort. Depending upon how powerful your machine is and how many rules you have defined, it may take

different amounts of time to respond to different packets. If traffic on your network is too high when Snort is working in NIDS mode, you may drop some packets and may not get a true real-time response.

2.2. Logging and Alerting System

Depending upon what the detection engine finds inside a packet, the packet may be used to log the activity or generate an alert. Logs are kept in simple text files, tcpdump-style files or some other form. All of the log files are stored under /var/log/ snort folder by default. You can use -l command line options to modify the location of generating logs and alerts. Many command line options discussed in the next chapter can modify the type and detail of information that is logged by the logging and alerting system [8].

3. Graylog

Graylog is a free and open source log management tool based on Java, Elasticsearch and MongoDB that can be used to collect, index and analyze any server log from a centralized location. You can easily monitor the SSH logins and unusual activity for debugging applications and logs using Graylog. Graylog provides a powerful query language, alerting abilities, a processing pipeline for data transformation and much more. You can extend the functionality of Graylog through a REST API and Add-ons [9].

Graylog has been successful in providing log management software because it was built for log management from the beginning. Software that stores and analyzes log data must have a very specific architecture to do it efficiently.

It is more than just a database or a full text search engine because it has to deal with both text data and metrics data on a time axis. Searches are always bound to a time frame (relative or absolute) and only going back into the past because future log data has not been written yet.

IV. ANALYSIS

By bringing those functions together, SIEM systems provide quicker identification, analysis and recovery of security events. They also allow compliance managers to confirm they are fulfilling an organisation's legal compliance requirements.

There are benefits to other areas of business as well: Improved productivity: SIEM systems allow trained security staff to move from unproductive, repetitive log file analysis to a more proactive role in the organisation. By analysing and correlating event logs from multiple devices, staff members are able to identify problems more easily. SIEM systems also provide a clear audit trail of events for compliance purposes.

Better handling of security breaches: IT staff can use

SIEM systems to produce rapid responses to security breach attempts as well as swift resolution of any problems. Consequently, SIEM systems minimise the cost of breaches and associated analysis and remediation, and from a compliance standpoint, offer clear processes for dealing with problems.

Optimisation of business processes: SIEM systems can provide an excellent overview of business processes and the use of business assets. This can allow organisations to make cost savings where assets are under-used, for example.

Business reporting: IT staff can use SIEM systems for reporting and analysis on a wide range of activity that can benefit several areas of an organisation, including the security team, IT management, finance, human resources and operations. Such information ranges from trend analysis to real-time and historical analysis of activity patterns.

Beside achievements, our system still has some downsides:

1. Don't have Cloud support for multi Company.
2. HIDS only support Windows and Linux platform.
3. Still cannot keeping logs for long term usages.

The solution can be extended and developed more to support Cloud database, make it easier for user to modify rules and storing logs in a long time for juridical purpose.

V. EXPERIMENTAL RESULT AND CONCLUSION

1. Test Results

Table 1. System tests and outputs

Test	Meaning - Result
Create, modify and delete files	Check integrity checksum alert - Passed
Login failed many times	Check alert when multiple login attempt failed - Passed
Modify registry keys	Check registry integrity checksum alert - Passed
Download files from Internet	Check alert when file be downloaded from Internet - Passed
Use OWASP ZAP	Check alert when website got attack - Passed
Perform Nmap scan – Port scanning technique	Check alert when port being scanned - Passed
Observe local traffic	Check traffic between local machines - Passed

2. Conclusion

The performance and usability of the solution are proved through extensive testing process. Compare

to others available similar approach, our system has several improvement:

The budget to implement our system is low since we only use open source program to build it . Admin users can modify alerting rules at ease to match organization's policy. Support user friendly enviroment.

ACKNOWLEDGEMENTS

First of all, we would like to express our gratitude toward our supervisor, Mr.Ngo Dang Ha An for helping us defining project scope as well as insightful suggestions and feedback to make our project improve more and more.

REFERENCES

- [1] Interpol.int. Crime-area cybercrime cybercrime. [online] available at: <https://www.interpol.int/crime-areas/cybercrime/cybercrime>
- [2] Alienvault.com. (2014). Everything you wanted to know about siem and log management but were afraid. [online] available at: <https://www.alienvault.com/blogs/security-essentials/everything-you-wanted-to-know-about-siem-and-log-management-but-were-afraid>
- [3] Securitymetrics.com. (2015). Importance of log management. [online] available at: <http://blog.securitymetrics.com/2015/08/importance-of-log>

management.html

[4] Symantec.com. (2012). Security information management vs security information and ev. [online] available at: <https://www.symantec.com/connect/articles/security-information-management-vs-security-event-management-vs-security-information-and-ev>

[5] Symantec.com. (2012). Security information management vs security information and ev. [online] available at: <https://www.symantec.com/connect/articles/security-information-management-vs-security-event-management-vs-security-information-and-ev>

[6] Rory bray, andrew hay, daniel cid, ossec host-based intrusion detection guide. 2008

[7] Rory bray, andrew hay, daniel cid, ossec host-based intrusion detection guide. 2008

[8] Rafeeq ur rehman, intrusion detection systems with snort. 2003

[9] Graylog.org. (2017) graylog - trusted full-featured log management. [online] available at: <https://www.graylog.org/>

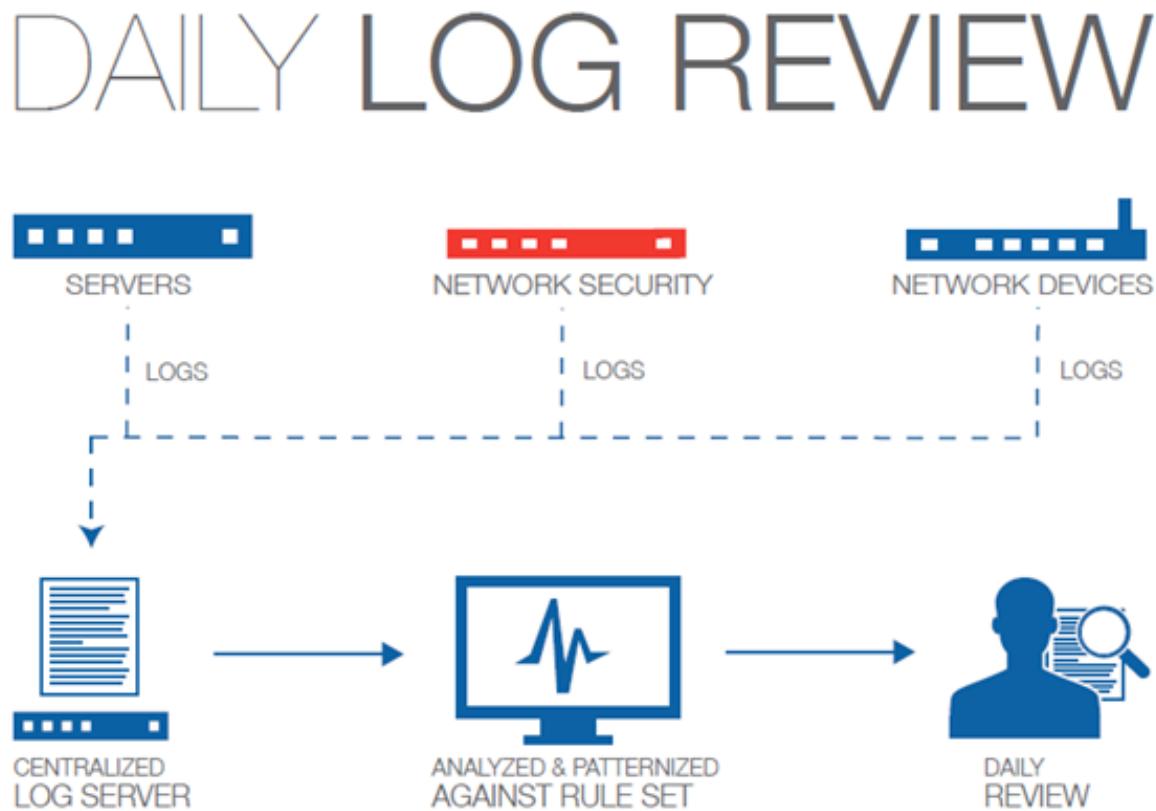


Figure 1: Process of simple Security Information Event Management

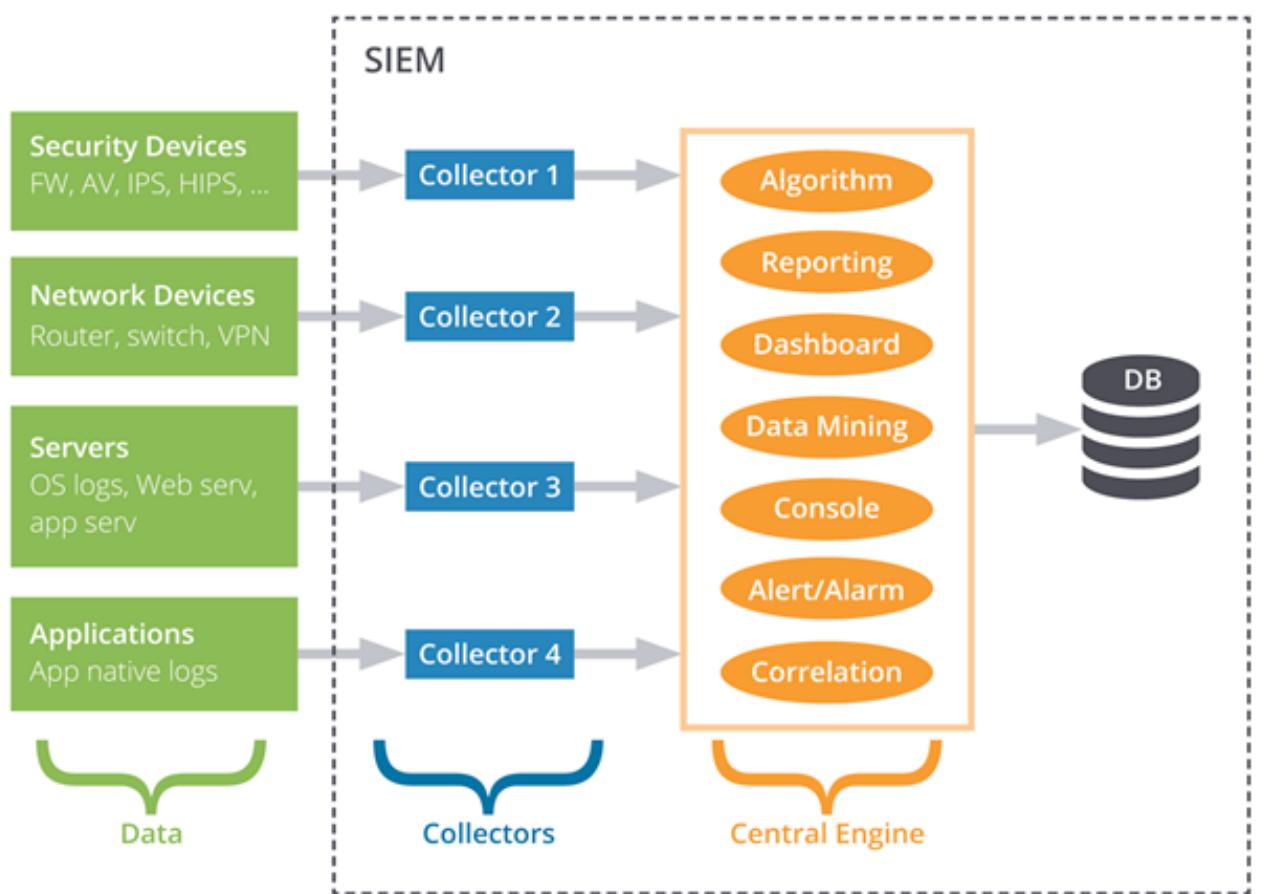


Figure 2: Detail Process of Security Information Event Management

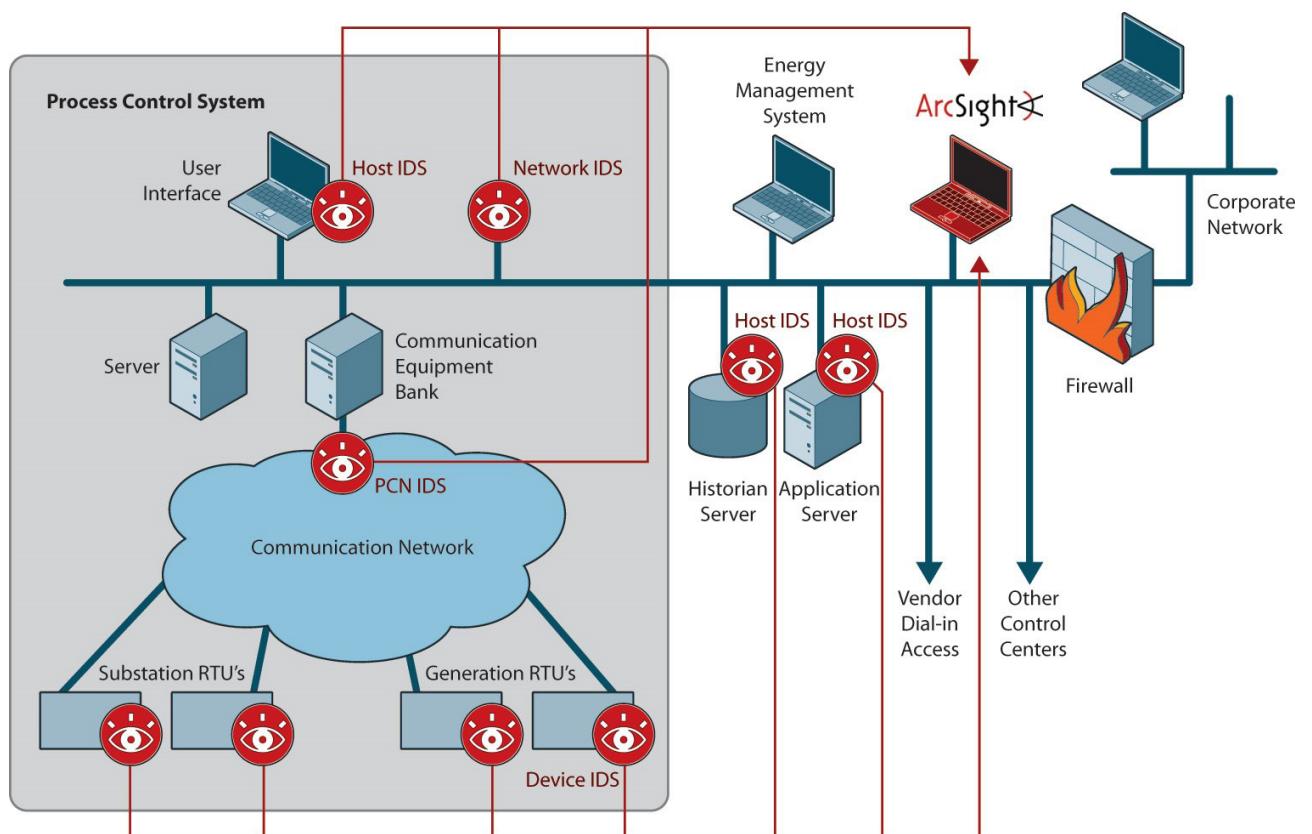
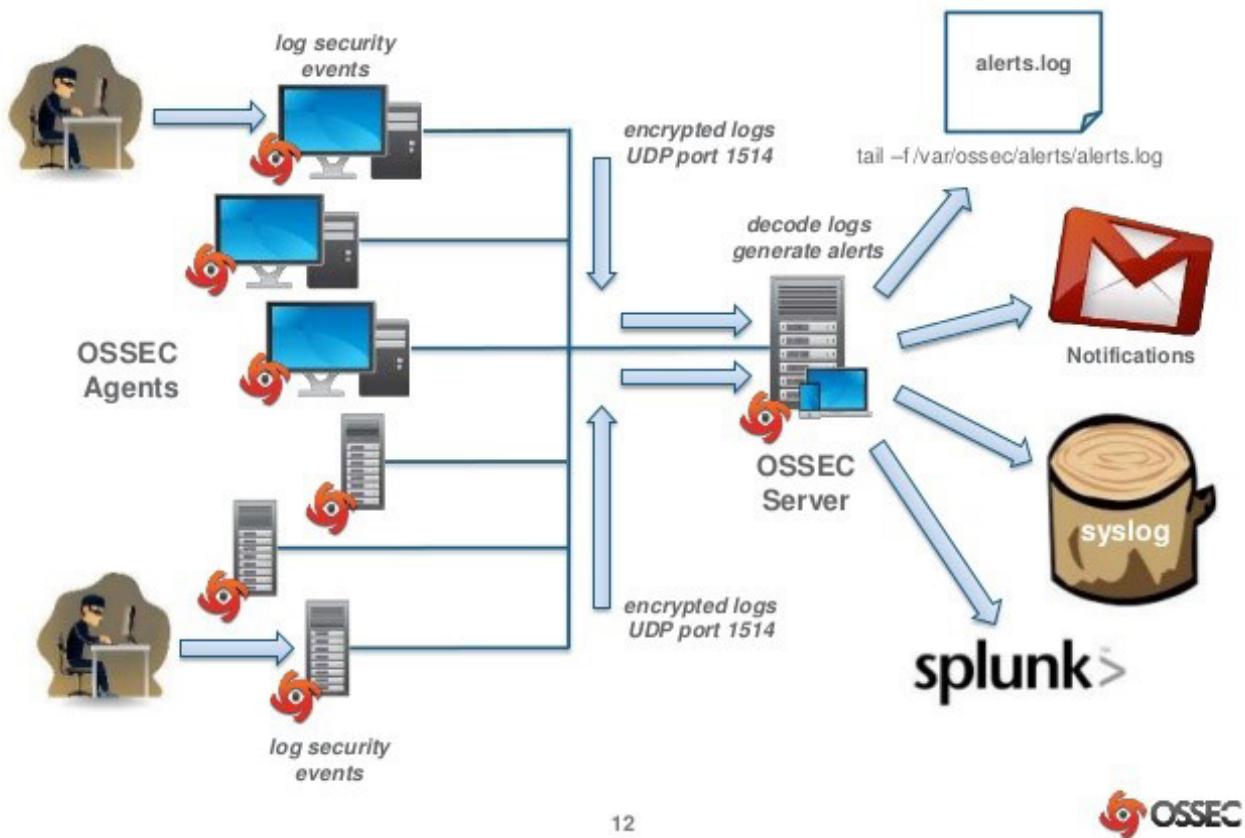


Figure 3: SIEM System

OSSEC In Action



12



Figure 4: Ossec In Action (Ossec is HIDS a part of SIEM System)

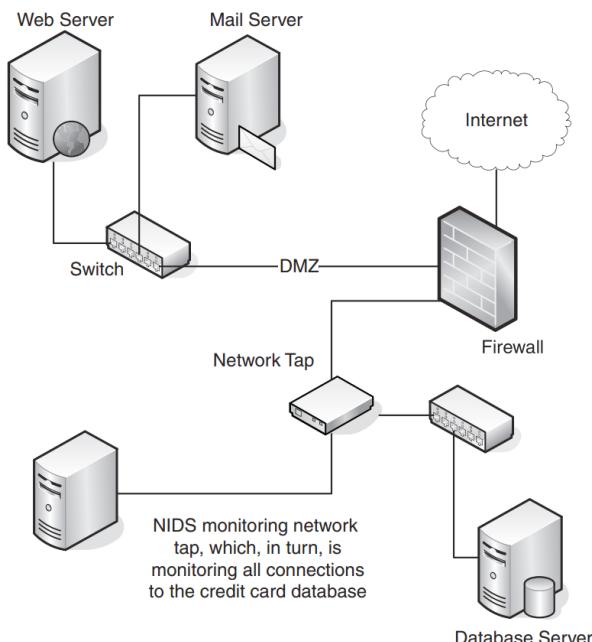


Figure 5: NIDS Deployment Monitoring Connections to an Internal Database

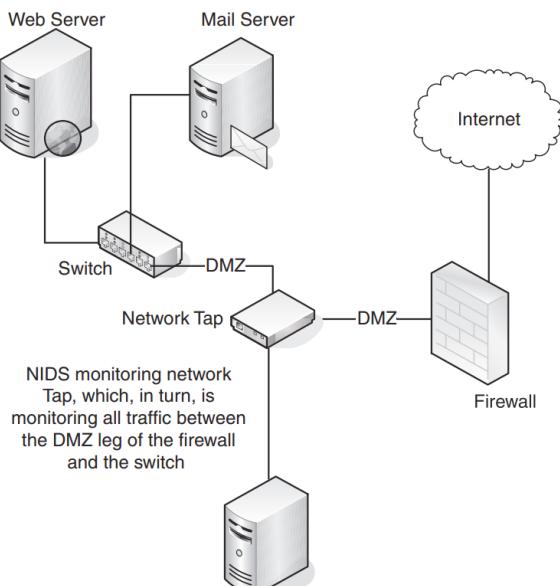


Figure 6: NIDS Monitoring the DMZ

Door Access and Payroll Monitoring System:

Deploying in small and medium company require a system to manage staff access

Nguyen Xuan Huy
FPT University
HCMC, Viet Nam

Bui Cong Khanh
FPT University
HCMC, Viet Nam

Bui Quang Vien
FPT University
HCMC, Viet Nam

Nguyen Xuan Huy
FPT University
HCMC, Viet Nam

Abstract

At small or medium sized companies or factories, door access and payroll monitoring system for employees are not optimized; they still use human resources as their main source. This causes some effects on the attendance process such as: bias towards one or several individuals or errors in the timekeeping process.

We will provide a system to limit that situations. Furthermore, we will provide managers with a more in-depth view of their employees' work processes. From there, managers can tailor the rules to improve employee productivity.

Keywords

Microcontroller; RFID; Access control; APMS

I. INTRODUCTION

Nowadays, door systems using RFID cards are familiar with us; companies use RFID cards to check attendance their employees, buildings use them to manage the population as well as the guests or parking lots use them to keep the car.

Therefore we chose this project: "Door access and payroll monitoring system" because of its practicality and popularity. With this project, we will design and implement a door system in combination with the web system to record employee arrival and exit times then calculate their salary based on data logged by the system.

II. PROBLEM AND SOLUTION PLAN

We have researched many of documents about current access systems, analyzing their strengths and weaknesses. We find out that most of these systems is hard to change when implement to other company or when the company change policy, the maintain cost is high and can only check the access data on computer that require to install their application.

After that, we find factors and give out solution to implement our manager system to web browser. Our three key points of our system are: can be easy to change in accordance with company policy, Implement the system on web that user can check their access information anywhere. Implement the payroll system into the access system.

At first, We build a system which allows user to access using RFID card, record the access data and send it to the payroll and manager system to calculate the salary. The system can be easy to configurations to company policy. The system will be implemented with the Door Access system and Payroll manager system.

Our solution contains a central controller which connects to the RFID readers, magnetic lock, LCD 16*2 monitor and communicate with local server. We have ... problems describe as below:

1. Control device with central controller

We develop our system to have main control board which contain Arduino Mega, Ethernet shield... and slave board for RFID reader modules which contain RFID readers, LCD16*2... .Main control board will connect with slave RFID Modules to get data access when card detect and send data about detected card to LCD 16*2 monitor.

2. Communicate I2C devices with main control board

Our systems will have at least 2 I2C device connect to it. To reduce the wiring and complexity of the main control board. We research and develop the I2C device that will connect to main controlled board and communicate with the same SDA and SCK line in I2C communication.



Figure 1: I2C Communication

3. Communicate SPI devices with main control board

Our systems will have at least 2 SPI device connect

to it. To reduces the wiring and complexity of the main control board. We research and develop the SPI device will connect to main control board and communicate with the same SCK,MISO, MOSI line in SPI communication.

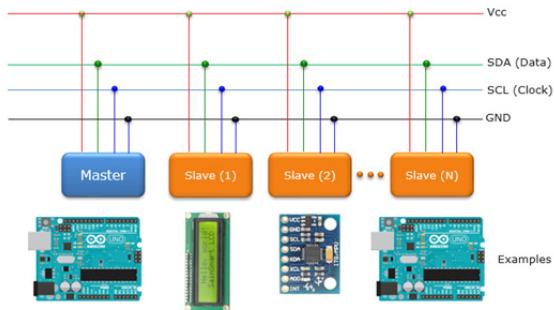


Figure 2: SPI Communication

4. Different voltage solution: IC L7805

We have used many device so the voltage for them is different from each other, magnetic lock using 12V, RFID reader using 3.3V and LCD screen using 5V . so we need to change the input voltage suitable for all device. So our solution is using power adapter 12 volts input and the IC L7805 (have output about 4.8v-5.2v and 1.5A) to to reduce the input voltage from 12 volts to 5 volts connect to arduino Mega to adjusting for RFID and LCD screen.

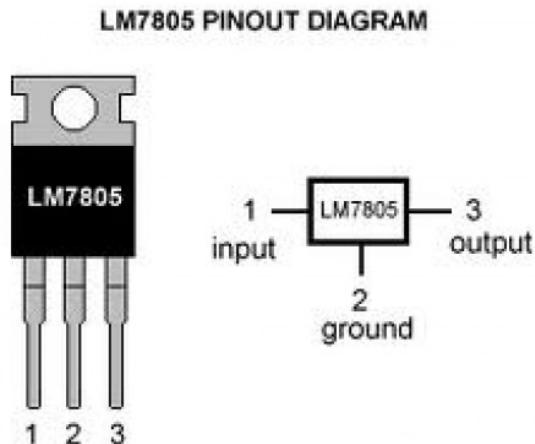


Figure 3: IC L7805

5. Control Solenoid Lock solution: Transistor NPN TIP31C (or other transistor more than 16v and 3A)

Because the Magnetic lock only have 2 connect wire (VCC and GND) so we need an electronic part that can control the magnetic lock when have access signal. Our solution is to use transistor NPN TIP31C to receive the signal from arduino mega and control the GND line of the magnetic lock as a digital switch.

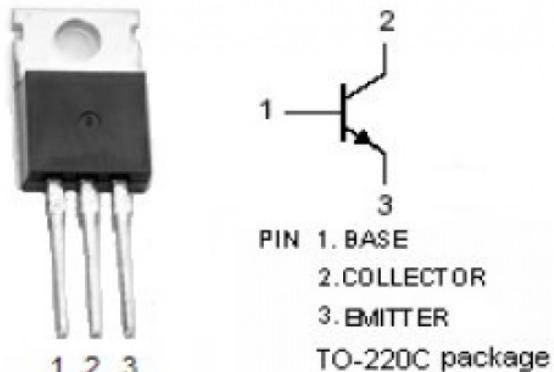


Figure 4: Transistor NPN TIP31C

6. Get and store server time even when the system restart and work offline: RTC – DS1307

Because Arduino Mega does not keep up the time when the system is restart so we need an extend module to store the time. Even when the system is off-line and restarted we still can get the time that we sync with the server at the latest connected to record when a card detect.

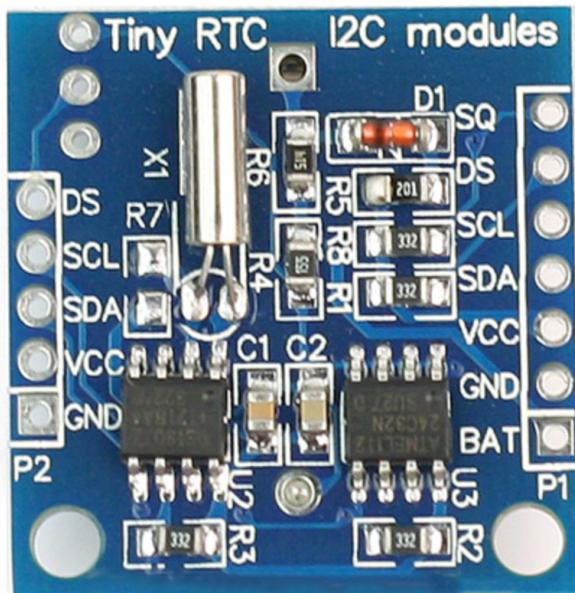


Figure 5: Module Real-time clock DS1307

7. Communicate with Local server and data stored when can not connect to local server: Ethernet shield

To communicate with the local server, we using Ethernet shield and it's also have Sdcard module embedded in it and using SPI communication to communication with the arduino mega.

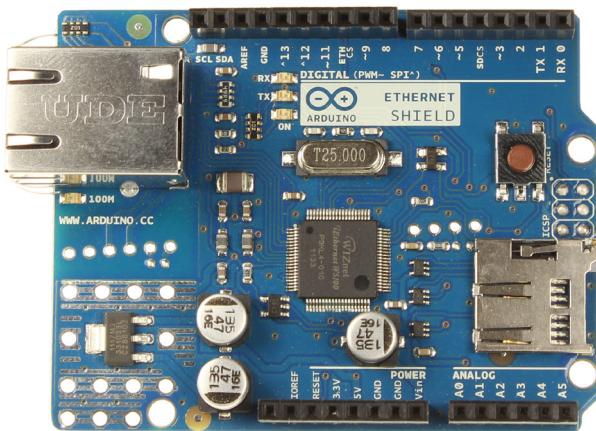


Figure 6: Module Ethernet Shield W5100

III. PLAN IMPLEMENTATION

1. Control device with central controller

We create and embedded control code to arduino that connect to main control board that can send data to local server and recive data or setup change from web application. At the time system turn on, the micro controller will connect to server to get data setup and sync server time to the RTC module. User with permission can use web app to configure the system hardware to suitable for company policy Component of main control board include: auto configurations, information of device. main control board communication with local though LAN via HTTP protocol. User uses an account provided from User have permission to create and give it to sign in into the Web App Manager.

Every modules connect with main control board has been configured by user with permission before the system is provided for company. Therefore, they can configure the web app and hardware modules as soon as login first time.

2. I2C and SPI communication Protocol with main board

Many of our devices communication with mainboard via I2C and SPI communication Protocol like Ethernet shield, RTC modules,2 RFID readerIf we use 1 I2C or SPI communication Protool port for each device that need these communication protocol to communicate with main board, that will cost alot of pin when we connect all of device together with the main board and the complexity of wiring will be alot more. So we decide to using 1 I2C and 1 SPI communication Protocol line to connect to all module that need to use these communication protocol into it.

With I2C communication protocal using 2 wire, 1 is SDA and 1 is SCK. Master device will sent clock frequence though SCK line and data (include device selected address) though SDA line to all

device connect to I2C line but only device with the same address include in the data will process data sending from master device (in case you have 2 or more device have the same address, you can use PCA9544A[1] module to select device you want to select when there more devices you using have the same address).

SPI communication protocol use 4 wires to communicate between master and slaves that are SCK, MOSI, MISO and SS (or CS). All slave devices can use the same SCK, MOSI and MISO line to connect to Master device. And the SS(CS) line of each device will connect to separate digital for each device to identify which device will get the data sent from master device.

3. Different voltage and control devices

Because we use a lot of devices, some of them will have different voltage then other devices so we will use 12 volt input to supply all our devices so we will use IC L7805 to lower voltage from 12v to 5v to supply for Arduino, LCD, buzzer and Arduino will lower voltage to 3.3V to supply for led and RFID reader. Beside we also need an electronic components that can act as an electric switch to control the magnetic lock, so we use transistor TIP31C to do it.

4. Synchronized and keep up time of the main board with the local server

Our system is an attendance checking system so we will need to synchronize time of server with main board and keeping the time up with server. RTC module DS1307 is suitable for save the synchronized time and keeping up the time so we can have the time that is synchronized when cards detected by the reader. We also need Ethernet shield to connect our main board with server and get the time from server.

5. Get and send data from main board to local server and stored when Ethernet offline

To send and recive data between main board and local server we using Ethernet sheild, because Ethernet sheild also include Sdcard module to stored card detect record and card user info to check cards ID is valid when mainboard cant communicate with the local server. When the main board reconnected to the local server, local server side can sent a command to the main board and request mainboard to synchronize data record of the time mainboard and server cant communicate.

IV. ANALYSIS

At first we implement our system on main board with Arduino Uno R3 board is the main processing unit which have Atmega328 8-bit MCU [2] with 32KB of flash memory and 2KB of SRAM*[3] but when we merge all control code, libraries of our

devices and compiling, it appears lacking of flash memory and SRAM making system run unstable, so we change our main processing unit to Arduino mega R3 which has ATmega2560 8-bit MCU with 256KB of flash memory and 8KB of SRAM. After doing system test we have some analyze and risk assessment below:

1. Main processing unit Arduino Mega R3

Because our main processing unit Arduino Mega R3[4] board have some modify to fit our main board. When we modify it may cause some problem to pins that we modify, this is unavoidable. There are some ways to reduce the change cause problem but it requires technical skills about soldering board. Because of lacking technical skills, after modifying the board we need to test every pin that we modified is work stable or not and list out the pin that have problem. To avoid using it in our main board.

2. RFID reader and RFID cards

We use RFID reader RC-522 modules which have working frequency at 13.56 MHz (HF) handles the the ISO14443A[6] frame and error detection. Cards have ISO14443 type A communicate via radio at 13.56MHz (HF). When a tag gets near the reader, it will absorb power spreaded by the reader and change the power absorbed to RF wave contain the id data in the card. The reader will get the RF wave contain the id data and decrypt it.

There are 2 common ways to put data on radio wave that is frequency modulation(FM) and amplitude modulation (AM). The RFID reader and RFID cards using amplitude-shift keying (ASK), which is a form of AM that represents the digital data as variations in the amplitude of a carrier wave.

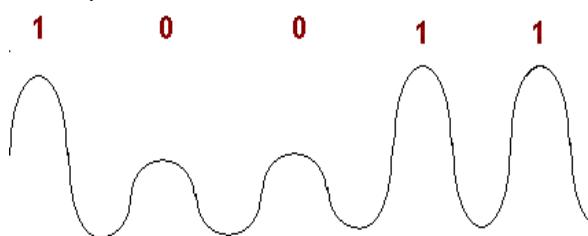


Figure 7: AM signal

To draw the amplitude modulation signal we use this algorithm

$$y(t) = A_c \cos(2\pi f_c t + 2\pi f_\Delta \int_0^t x_m(t) dt)$$

A_c : the amplitude of carrier wave.

f_c : the frequency of the carrier wave.

f_Δ : the frequency offset.

$x_m(t)$: The signal need to transmit.

Which is a sinusoidal continuous signal with

frequency f_m and amplitude A_m have integral is

$$\int_0^t x_m(t) dt = \frac{A_m \cos(2\pi f_m t)}{2\pi f_m}$$

ASK modulation is a method that if the bit sent is "1" then the carrier signal will be transmitted, else the signal value "0" will be transmitted.

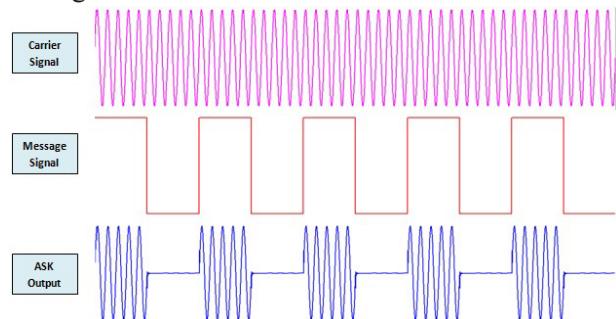


Figure 8: ASK Signal output

3. Control access using Solenoid lock LY-03

Our system is a system that requires a way to check people in and out of a zone, so we use Solenoid lock LY-03 to control access. The Solenoid Lock provide a good response time, below 1 second, but we will need to change to other module if apply it to real product. Because the Solenoid lock LY-03 heats up if we provide continuous electricity through it over 10 seconds.

4. Power control and heat generated by IC 7805

We use IC L7805 to reduce voltage from our 12V power source which is used to activate Solenoid lock (12 V and 0.8A) to 5V to supply our Arduino Mega (30mA max at 5V for MCU and other components on board, 20mA at 5V for each of 11 digital pins used), 2 LCD 1605 (30mA max at 5V), 2 RC-522 (26mA max at 3.3V), 2 Buzzer 5V (25mA at 5V), DS1307 RTC-Module (has separate power source), Ethernet shield (max 183mA at 3.3V) and FC-51 Infrared Obstacle Avoidance Proximity Sensors Module (23mA at 3.3V). So our system will draw max 530.28mA at 5V from IC L7805 and generate about 2.7W of heat. And total consumption of our system is about 1331mA at 12V so we need a power source that can provide more than 1.7A at 12V (80% efficiency) and a heatsink [5] for IC 7805.

5. Communicate via LAN and local stored by Ethernet shield W5100

We use module Ethernet shield W5100 which has maximum current consumption is 183mA at 3.3V, about 0.6W power consumption. So the chip may generate some heat on it even when idle, heatsink is recommended when using this module.

V. EXPERIMENTAL RESULT AND CONCLUSION

With our practical experience, we have learned and tested with each devices. After that, we integrate all of devices together, but we have some problem when integrate its such as: conflict between SPI communication devices, current of power supply, and business logic of our system.

So, to have a finished product we must have a test step with criteria are:

- 90% of the test cases must pass
- 100% of test cases about hardware module must pass
- All test cases dealing with critical functionality must pass
- All medium and high severity defects must be fixed
- Test coverage must be at least 90%

After done testing step above, the final product is working as our expectation.

In conclusion, we create a system that can communication to local to get data when condition meet or getting data from local when change applied. Acknowledgment: First and foremost, we would like to express our greatest gratitude to Mr.Nguyen Duc Loi our Advisor for his support and guidance during this project, especially in fulfilling project requirements and improving project's quality.

In addition, we would like to send our special thanks to all FPT University Lecturers who have taught and guided us from the very beginning. This has been a great instruction in this strenuous journey.

REFERENCES

- [1] PCA9544A
<http://www.dsscircuits.com/sale/product/dssc0107>
- [2] Atmega328 8-bit MCU
<https://en.wikipedia.org/wiki/Microcontroller>
- [3] SRAM
https://en.wikipedia.org/wiki/Static_random-access_memory
- [4] Arduino Mega 2560: <http://hshop.vn/products/arduino-mega-2560r3>
- [5] Heatsink
https://vi.aliexpress.com/store/product/10pcs-lot-Aluminum-Heatsink-TO-220-20-15-10MM-TO220-Cooling-Heat-Sink-For-L7805-LM337/1725790_32332723712.html
- [6] ISO14443A
https://en.wikipedia.org/wiki/ISO/IEC_14443

Electronic Devices B2B Website

Nguyen Hoang Phuc

FPT University

HCMC, Viet Nam

Nguyen Thi Thai Han

FPT University

HCMC, Viet Nam

Nguyen Van Bach

FPT University

HCMC, Viet Nam

Phan Nhat Minh

FPT University

HCMC, Viet Nam

Abstract

Nowadays, thanks to the significant development of numerous types of business and the E-commerce business model – B2B (Business-to-business) has emerged. B2B is a model where transactions are carried out between businesses including suppliers, manufacturers, as well as small and medium enterprises. Transaction and contracts can be conducted directly and immediately through the Internet, eliminating the need to meet in person.

Keywords

B2B Website; Electronic Device B2B; E-commerce; B2B marketplace; B2B system

I. INTRODUCTION

Electronic Devices B2B Website was born based on B2B – E-commerce business model. Those websites provide many features supporting businesses such as offering products, seeking trustworthy partners, ordering supply, signing contracts and purchasing via payment systems. The success of the website depends on 5 main functions: The website offers ordering online – Make Order [1] – This website allows users to search for products based on categories and brands. Then the users can order products via this system. Finally, the products will be delivered to the given addresses by suppliers. The website plays an intermediary role in order to ensure secure and smooth transactions.

The website predicts the products that you may like – Recommended List [2] – The website uses a product-suggesting function based on previous behaviors and ratings. Then, the website recommends those products meeting users' criteria in choosing.

Instead of offering a communal service and assisting in selling, we concentrate on the behaviors and search criteria of users. The system provides services in the fastest and most effective way for users' convenience.

So as to achieve that, the suggesting feature has been enhanced so that customers can get more flexible and wiser suggestions. The system uses users' product reviews and feedback as inputs, calculates them, and gradually make suggestions depending on evaluation results.

The website allows creating bidding contracts – Create Tendering [3] – The website assists users in seeking for trustworthy suppliers through a function which creates bids and assesses contractors. The supplier will provide the best criteria for winning the bid. Consequently, users can make better assessment and build trust among prospective partners.

The website offers online negotiation among users – Negotiate [3] – The website supports users in real-time interacting to supplier and through an online messaging system aiming to maximize the success possibilities of transactions. When a buyer has special requirements for an order, he or she can contact the supplier directly through this function.

The website creates group-buying deals – Create Group-buying [4] – The website allows users (suppliers) create group-buying deals so that customers can get discounts. Customers can buy products at a low price. Besides, suppliers can also make more sales as well as expand the sales market through this measure.

Apart from improving search engines, the system provides for a reasonable suggestion. In order to enhance convenience, the system displays the products considered to be suitable for users. Users can approach many more similar and related products. Also, some online purchasing methods via payment gateways are integrated in the system to ensure security and convenience for users.

In Section 2 we formalize the problem and describe our approach to solving it. Section 3 describes the data structures that need to be set up and the algorithms. We analyze our algorithm in Section 4. Experimental results and the software used are described in the concluding Section 5.

II. PROBLEM AND SOLUTION PLAN

We divided into the following main functions:

► Recommended System:

- **Problems:** Nowadays, there is a large number of electronic devices created. It is quite difficult for

Supervisor: Lai Duc Hung, M.E

(Lecturer of IT Department, HCMC FPT University)

Defended topic in front of the scientific council of HCMC

FPT University on December 14th, 2017

Accepted post on December 26th, 2017

users to choose appropriate products and similar ones. The set goal is to suggest effectively and appropriately. By that way, the system would help customers save time and money

- **Solution:** The approach to solving this problem is to give some recommended product list that help users find product quick and easy:

○ User-User Collaborative Filtering:

Recommendations are based on analysis of the similarities between users through evaluation data from them. The data that the system uses is the user interaction performed, or is the user's evaluation of the products viewed. Based on that, the system predicts a user's interest for an item, uses the interest of similar users and recommends some products that similar users also see, which are satisfactory to contemporary users.

○ Item-Item Collaborative Filtering:

This recommendation based on the similarity between products calculated using the user's reviews of those products. Based on that, users will be offered some products that are most similar to the product the user is viewing.

► Real-Time Interaction:

- **Problems:** The trouble with online shopping is that many customers are disconnected from someone who can answer their questions in real time with precision. Rather than watching potential customers click away from their e-commerce sites, many businesses have been adding live chat support. As it turns out, live chat has the ability to provide the convenient answers that customers want, while also adding significant benefits to the staff and bottom line of companies.

- **Solution:** WebSockets are a bi-directional, persistent connection from a web browser to a server. Once a WebSocket connection is established the connection and stays open until the client or server decides to close this connection. With this open connection, the client or server can send a message at any given time to the other. This makes web programming entirely event driven, not just user initiated. Furthermore, at this time, a single running server application is aware of all connections, allowing you to communicate with any number of open connections at any given time. Therefore, this is the best solution to solve the mentioned problem.

III. PLAN IMPLEMENTATION

1. System Architecture

- Our system is developed based on MVC architectural style. We choose this architecture because of the following advantages:
 - We can organize the code better for maintainability, extensibility, and reusability

- Business logic developers can build the classes, while the UI developers can involve in designing UI screens simultaneously, resulting the interdependency issues and time conservation.

2. Database

- MySQL: We choose MySQL because some reasons:

- Data Security: MySQL is globally renowned for being the most secure and reliable database management. The data security and support for transactional processing that accompany the recent version of MySQL can greatly benefit any business especially if it is an e-commerce business that involves frequent money transfers.
- On-Demand Scalability: On-demand flexibility is the star feature of MySQL. This open source solution allows complete customization to e-commerce businesses with unique database server requirements.
- High Performance: MySQL features a distinct storage-engine framework that facilities system administrators to configure the MySQL database server for a flawless performance. Whether it is an e-commerce website that receives a million queries every single day or a high-speed transactional processing system.

In order to solve above problems, we apply technologies below, combine them into the algorithms which is described below:

- User-User Collaborative Filtering
- Get the review data of users
- Calculate the normalized matrix users with items.
- Calculate the similarity between these users.
- If users have rated any product, the system will predict products you may like based on similar users.
- Calculate similarity by formula:

$$\text{similarity} = \cos \theta = \frac{\sum_{i=1}^{n \in A \cap B} x_{A_i} \times x_{B_i}}{\sqrt{\sum_{i=0}^{n \in A} x_{A_i}^2} \times \sqrt{\sum_{i=0}^{n \in B} x_{B_i}^2}}$$

Where:

- A: List of user A's level of interest on each product.
- B: List of user B's level of interest on each product.
- Calculate the level of user interest on each product by formula:

$$\hat{y}_{i,u} = \frac{\sum_{u_j \in N(u,i)} \bar{y}_{i,u_j} \text{sim}(u, u_j)}{\sum_{u_j \in N(u,i)} |\text{sim}(u, u_j)|}$$

Where:

- \bar{y}_{i,u_j} : User u_j 's level of interest on product i .
- $\text{sim}(u, u_j)$: Correlation of user A and user B.
- $N(u,i)$: List of the interest level of users has the most similarity.
- Calculate the level of user interest on all products.

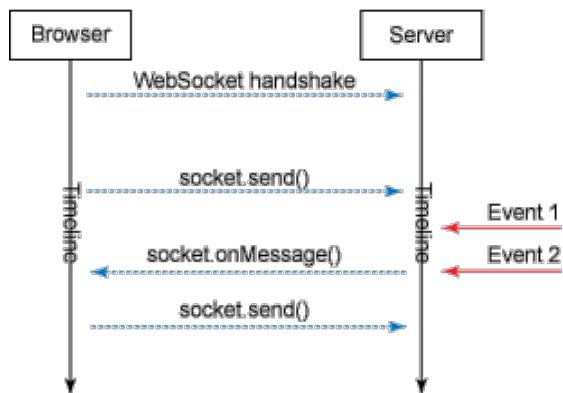
- Sort above list by descending and recommend some products that user may like.
- The complexity of this algorithm is N^{-2} .
- Item-Item Collaborative Filtering
- Get the review data of users
- Calculate the normalized matrix users with items.
- Calculate the similarity between these items.
- Calculate similarity by formula:

$$\text{similarity} = \cos \theta = \frac{\sum_{i=1}^{n \in A \cap B} x_{A_i} \times x_{B_i}}{\sqrt{\sum_{i=0}^{n \in A} x_{A_i}^2} \times \sqrt{\sum_{i=0}^{n \in B} x_{B_i}^2}}$$

Where:

- A: List of all user's level of interest on product A.
 B: List of all user's level of interest on product B.
- Sort above list by descending.
 - Recommend some products which correlate with the product that user is viewing.
 - The complexity of this algorithm is N^{-2} .

■ Real-Time Interaction



We use the following 3 events:

- onopen: When websocket is opened
 onmessage: When receiving a message
 onclose: When websocket is closed

To establish a WebSocket connection, the client sends a WebSocket handshake request, for which the server returns a WebSocket handshake response, as shown in the example below.

- Client request:
 GET /mychat HTTP/1.1
 Host: server.example.com
 Upgrade: websocket
 Connection: Upgrade
 Sec-WebSocket-Key: x3JJHMbDL1EzLkh9GBhXDw==
 Sec-WebSocket-Protocol: chat
 Sec-WebSocket-Version: 13
 Origin: http://example.com
- Server response:
 HTTP/1.1 101 Switching Protocols
 Upgrade: websocket
 Connection: Upgrade
 Sec-WebSocket-Accept:

HSmrco0sMIYUkAGmm5OPpG2HaGWk=

Sec-WebSocket-Protocol: chat

To confirm the connection, the client sends a Sec-WebSocket-Key encrypted value from the Base64 to the server. Then the server will perform:

– Append the fixed string “258EAFA5-E914-47DA-95CA-C5AB0DC85B11” to Sec-WebSocket-Key to get the new string “x3JJHMbDL1EzLkh9GBhXDw==258EAFA5-E914-47DA-95CA-C5AB0DC85B11”.

– Apply encrypted SHA-1 string on to be “1d29ab734b0c9585240069a6e4e3e91b61da1969”.

– Encrypt the result received by Base64 to be “HSmrco0sMIYUkAGmm5OPpG2HaGWk==”.

– Sending the response to the client with the Sec-WebSocket-Accept value is the resultant result set. The client will check the status code (must be 101) and Sec-WebSocket-Accept to see if it matches the expected results and make the connection.

IV. ANALYSIS

In this section, we analyse the performance and functionality of 2 algorithms which mentioned below:

- User-based nearest Collaborative Filtering
 The complexity of this algorithm is N^2 .
 - We execute User based nearest Collaborative Filtering.
 - Data of user's review updated daily.
 - After execution, we receive response depend on amount of input data.
- Item-Item Collaborative Filtering
 The complexity of this algorithm is N^2 .
 - We execute Item-based Collaborative Filtering and receive response depend on amount of input data.

V. EXPERIMENTAL RESULTS AND CONCLUSION

ACKNOWLEDGMENT

We would like to thank to Mr. Lai Duc Hung for helping us define project scope and improving the functionality.

REFERENCES

- [1] <https://www.gocom.vn/>
- [2] <https://www.alibaba.com/>
- [3] https://en.wikipedia.org/wiki/Collaborative_filtering
- [4] <https://machinelearningcoban.com/2017/05/24/collaborativefiltering/>
- [5] <https://en.wikipedia.org/wiki/Business-to-business>
- [6] <http://socketo.me/docs/>
- [7] <http://blog.rikkeisoft.com/seminar-gioi-thieu/>

ve-websocket-va-node-js/
[8] [https://sparktutorials.github.io/2015/11/08/
spark-websocket-chat.html](https://sparktutorials.github.io/2015/11/08/spark-websocket-chat.html)

Football Field Reservation System

Mai Minh Quy
FPT University
HCMC, Viet Nam

Truong Huu Thanh
FPT University
HCMC, Viet Nam

Phan Minh Huan
FPT University
HCMC, Viet Nam

Pham Trung Hieu
FPT University
HCMC, Viet Nam

Abstract

Football is becoming popular with the increase in the number of players and field owners. Services, however, are not good enough to meet the growing demand. Usually, people who want to reserve a football field have to go to the field to reserve, or call the field owner. Otherwise, they have to give up because they just know few familiar football fields. For the field owners, they cannot find the channel to advertise their fields to potential users. That is the reason why we create Football Field Reservation System to resolve these problems, and provide some additional services for users and field owners.

Keywords

Reserve field; separate time slot; optimize free time; matching opponent request; location; haversine formula; constants distance.

I. INTRODUCTION

When user wants to reserve a field, user has to go to field to ask field owner if there are available fields. If the user wants to find an opponent, user must also request the field owner to find an opponent team. If field owner finds an opponent team, then the field will be reserved. Opponent team is found not based on any criteria excepts they have same free time. So, user cannot find the suitable opponent for level. For competitive sports, the difference in level will make the game becomes boring for both. It leads to needs to simplify the field reservation process and support to find opponents more suitable.

Solutions that the Football Field Reservation System provides is that it allows field owners and users to communicate in a quicker and more convenient way. In particular, the main purpose of the System is that field owners can manage their fields more effectively, while users can reserve fields online and find the same level opponents based on rating points. Rating points are calculated by ratings and comments of users after a matching match has finished.

In section 2, we will access the problem and describe our approach to solve it. Section 3 is the section we present detail about what our system does and algorithms for optimizing reservation time and calculate distance between two GPS coordinates on map. In section 4, we analyze solution. Section 5 presents the experimental results and our conclusion.

II. PROBLEM AND SOLUTION PLAN

For the field owner, conflicts and missing are likely to occur when they process reservation requests on paper. At idle hours, field owners are hardly able to find players or very little although they can reduce reservation field fee to a very low level. Players reserve field by calling and pay reservation fee after playing so the reliability is not high. The field owner must face the bad players, who reserves field but does not come to.

For the user, some very common problems occur: players spend too much time looking field have free time during peak hours, hard to find opponents, etc. The Football Field Reservation System provides services to 2 types of customers:

1. To the field owner, the system helps with field management, time management and price management of football fields. Field owner can launch promotion schemes to attract players during idle time.
2. The system helps user find field. The opponent search function allows the user to find the opponent that suits him/her most. If no suitable opponent is found, then the user can leave the matching request to other users who can find him/her.

III. PLAN IMPLEMENTATION

When user reserves field:

The system helps the user find the nearby fields. The system locates the coordinates of the user through GPS (Global Positioning System) and compares theirs with the coordinates of available fields in the database. The user can also choose search distance and new locations (via Google API). The System will inform user when the field is available. Then user will select the time frame and reserve the field. However, if many users are looking for fields at the same time, system must calculate between user's position with each field's position in database

Supervisor: Kieu Trong Khanh, M.E
(Head of ITS Department, HCMC FPT University)
Defended topic in front of the scientific council of HCMC
FPT University on December 13rd, 2017
Accepted post on December 26th, 2017

to return field list with distance are the closest. To improve the performance, the system uses some constants to support calculating distance between two points with difference of latitude and longitude. By calculating the difference between longitude and latitude, system will query to get the most suitable fields instead of get all field to calculate.

► Calculate Deviation Constants To Get Quickly Points Around With Expected Distance:

The situation is: when system finds fields around 1 location A (latitude1, longitude1), with k(km) radius. If location B (latitude2, longitude2) is within a radius of k(km) from A, the precondition is:

$$\begin{cases} \text{latitude}_1 - x < \text{latitude}_2 < \text{latitude}_1 + x \\ \text{longitude}_1 - y < \text{longitude}_2 < \text{longitude}_1 + y \end{cases}$$

x, y are deviation constants.

The main purpose is we calculate the approximately x, y then input to the system. Based on x, y, system can quickly find out which fields is the most suitable. X, y constants help system early removal fields cannot afford.

Calculating the distance between 2 points is one of the most used functions in our system (user reserves field, system finds opponent, system matches 2 users).

Calculating the distance between 2 points takes so many steps.

We calculate x, y as 2 constants. System based on x, y can calculate faster.

For any two points on a sphere, the haversine of the central angle between them is given by:

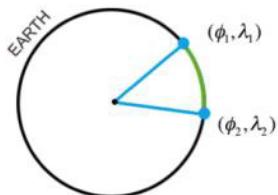
$$\text{hav}\left(\frac{d}{r}\right) = \text{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\text{hav}(\lambda_2 - \lambda_1)$$

hav is the Haversin function:

$$\text{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right)$$

- d is the distance between the two points
- r is the radius of the sphere,
- φ_1, φ_2 : latitude of point 1 and latitude of point 2, in radians
- λ_1, λ_2 : longitude of point 1 and longitude of point 2, in radians

$$\text{haversine}\left(\frac{d}{r}\right) = \text{haversine}(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\text{haversine}(\lambda_2 - \lambda_1)$$



Note that φ and λ can be converted from radians to degrees by multiplying by as usual.

A (latitude1, longitude1), B (latitude2, longitude2)
The latitude of Ho Chi Minh City, Vietnam is 10.762622, and the longitude is 106.660172.

- Let latitude1 = latitude2 = 10.76, d=5(km), r=6.367.449(m)

- longitude2 = longitude1±0.0455 (in decimal)

- Let longitude1 = longitude2, d=5(km), r=6.367.449(m)

- latitude2 = latitude1±0.045 (in decimal)

After testing, the erroneous is less than 5%.

A(10.8547079,106.6257702) (Google maps define as Quang Trung Software Park)

B (latitude, longitude) has longitude the same as A, latitude as A+0.0455

B(10.9002079,106.6257702)

The distance between A and B is 5.04km (using google maps calculator)

C (latitude, longitude) has latitude the same as A, longitude as A+0.045

C (10.8547079, 106.6707702)

The distance between A and C is 4.98km (using google maps calculator).

Complexity: The complexity of this algorithm is O(n).

When users find opponent:

The search opponent function allows user to find suitable opponents. The system will rely on the user's rank, distance between two users and their favorite fields list to make a list of suitable fields for both. If there are no suitable opponents, the user can leave the matching request for other users who can find him/her later.

Set favorite field and blacklist functions make the matching system more accurate and preferable to users. After completing a match, two users will rate the match, which enables the system to calculate 2 users' rank more accurately. To encourage 2 users to rate, the system will reward each user who gives correct/ honest rating bonus points for exchanging voucher.

System constructs calculating ranking scores method the same as FIFA (International Federation of Association Football). System can accurately assess the player's ability to create a decentralized environment that will make the sports more exciting.

Managing field:

Using the web application, field owner can create fields, time enable for each day of the week and set price for each time frame. The system supports field owner to optimize peak time periods. In optimizes time frame, the user must reserve the whole-time frame set by the field owner. In order to facilitate the transference to web application, the system allows field owner to reserve a period of time, provided

that no other users have previously reserved it. The online payment system also allows payment become quick and easy, which minimizes the risk to the field owner. Field owner can launch promotion schemes to attract players during off-peak times. The system also sets minimum price for every time frame and every location to prevent anti-dumping and create a healthy business environment for field owners.

► Optimize time when suggest field to reserve:

The field management system is based on vacant time frames of the field, so that the System can know which field owner can satisfy the user's request. If $H(0, \dots, k, \dots, 23)$ is the set of active fields for one hour from k to $k + 1$ ($H(k)$, the number of active fields between k and $k + 1$ hour).

When a user reserves the field from x to y hour(s). The system will check $H(k)$ with $x \leq k \leq y-1$ with all $H(k) > 0$, there exists at least 1 vacant field from x to y .

Let $F(1, \dots, i, n)$ is the set of fields of a field owner. User A reserve field from x to y hour(s).

Step 1: The System will check $H(k)$ with $x \leq k \leq y-1$ with all $H(k) > 0$, there exists at least 1 vacant field from x to y .

Step 2: Suppose $F(i)$ is the field that Field owner choose for user A.

For $(j = x, j = j + 1, j \leq y)$

If $(F(i))$ is free at j hour) then

User can reserve this field as normal till j hour

If $(F(i))$ is reserved by other user) then

Field owner find other field $F(ii)$ which has free time at j hour

Swap all reserve request of $F(i)$ and $F(ii)$ from j hour
After swap $F(i)$ is free at j hour

After the loop finish, field $F(i)$ has free time from x hour(s) to y hour(s). So that user can reserve at field $F(i)$.

Based on that, our system just has to manage free time of field owner. We don't need to know exactly what field User reserve, just need to know if user can reserve field or not.

● Example:

Two fields of field owner "Cây Trâm":

- Field 1: 06AM to 22PM
- Field 2: 06AM to 22PM

● Reserve request:

- User A reserved Field 1, time frame is: 06AM to 08AM
- User B reserved Field 2, time frame is: 08AM to 10AM

New reserve request:

- User C reserves time frame 07AM to 09AM

Field owner set field for User:

Field / Time	06h - 07h	07h - 08h	08h - 09h	09h - 10h	10h - 11h	11h - 12h
Field 1	User A	User A	0	0	0	0
Field 2	0	0	User B	User B	0	0

After system apply algorithm:

Field	06h – 07h	07h – 08h	08h – 09h	09h – 10h	10 h – 11h	11h – 12h
Field 1	User A	User A	User B	User B	0	0
Field 2	0	User C	User C	0	0	0

IV. ANALYSIS

After 2 users create matching request, if 2 requests are matched, system will suggest list of fields for user. The system not only match users based on time but also rely on the user's rank, distance between two users and the list of their favorite fields to make a list of fields suitable for both. Set favorite field and blacklist functions make the matching system more accurate and preferable to users. After completing a match, two users will rate the match, which enables the system to calculate 2 users' rank more accurately. The favorite field has high priority on the suitable list of fields. If a user is in the blacklist of another user, they cannot match each other.

When a large number of user reserve field, the field's timeline may be broken into fragments. Manage time fund make the system has ability to avoid that. System can decide what field suits the most.

V. EXPERIMENTAL RESULTS AND CONCLUSION

A. Test Result

Test No.	Average page load time (second)
1	0.85
2	1.1
3	0.7
4	0.9
5	1.3
	Average: 0.97 (seconds)

B. Conclusion

The system suggests more accurate when there are many field owners. The system does not take too much time to match 2 users. The system performs more slowly when multiple requests are made at the same time, but acceptable.

ACKNOWLEDGMENT

We would like to thank to Mr. Kieu Trong Khanh for

helping us defining project scope and improving the functionality.

REFERENCES

- [1] Android Developer Guide - Application Fundamentals:
<http://developer.android.com/guide/components/fundamentals.html>
- [2] <http://www.agiledata.org/essays/evolutionaryDevelopment.html>

Hacking and Security Techniques on Web Applications

Ho Bao Tien
FPT University
HCMC, Viet Nam

Hoang Quoc Hung
FPT University
HCMC, Viet Nam

Lam Ngo Quyen
FPT University
HCMC, Viet Nam

Huynh Thanh Trung
FPT University
HCMC, Viet Nam

Abstract

In recent years, along with the development of the internet, Web Applications are becoming more and more widely used and popular. Web Applications appear on much of today's online business, such as banking, shopping, social networking, and various governmental activities. As a result, there are more and more cyber attacks whose target is the web applications. Recognizing the serious problem, firstly, we took an analysis on several hacking techniques based on OWASP Top 10 document. Secondly, we built a Web Application Firewall (WAF) that applies "Data Mining" technology to detect and prevent hacker attacks from Web environment.

Keywords

Hacking techniques; WAF; Data mining; Decision Tree.

I. INTRODUCTION

According to statistics, the numbers of security vulnerabilities that are being found today are much higher in Applications than in operating systems. This means that the attacks aimed at Web Applications are exploiting vulnerabilities at the Application level and not at the transport or network level like common attacks from the past. At the same time, quantity and impact of security vulnerabilities in such Applications has grown as well. Many transactions are performed online with various kinds of Web Applications. Almost in all of them user is authenticated before providing access to backend database for storing all the information. [1] There are known vulnerabilities that simple programming practices can reduce. However, there are some vulnerabilities that are difficult to detect, when a vulnerability is exploited without being exposed by any mechanism, it can cause unpredictable threats and impact not only with the web app but also for the system.

In this paper, we will take a closer look at the attack techniques on the web app. Some may be very familiar and common, but others are still very new and difficult to detect. However, these attacks are dangerous and have a certain impact on the web application system, the data and information contained in the web app, which can sometimes be critical and important data. And more or less it will affect the reputation of the company if attacked. After listing attack techniques, it is necessary to introduce a security technique. The solution we choose here is building a Web Application Firewall (WAF) to detect and prevent hacker attacks on the web app. Although this is a common practice and widely used on the web app, it only detects attacks based on certain signs (such as Regular expression, etc.) and not capable of detecting new attacks. However, with the improved method that we used, the application of "data mining" technology will enable the system to automatically identify and analyze new types of attacks based on existing data.

II. PROBLEM AND SOLUTION PLAN

There are many attack techniques that a hacker can use to attack a Web Application. Each of which has several attack methods, so within the scope of this paper, we will only list top 10 Application security risks usually found on Web Application, reference to The Open Web Application Security Project by OWASP [2]:

1. Injection: Injection flaws, such as SQL, OS, XXE, and LDAP injection occur when untrusted data is sent to an interpreter as part of a command or query. The attacker's hostile data can trick the interpreter into executing unintended commands or accessing data without proper authorization.

2. Broken Authentication and Session Management: Application functions related to authentication and session management are often implemented incorrectly, allowing attackers to compromise passwords, keys, or session tokens, or to exploit other implementation flaws to assume other users' identities (temporarily or permanently).

3. Cross-Site Scripting (XSS): XSS flaws occur whenever an application includes untrusted data in a new web page without proper validation or escaping,

Supervisor: Nguyen Sieu Dang, M.E
(Lecturer of IT Department, HCMC FPT University)
Defended topic in front of the scientific council of HCMC
FPT University on December 15th, 2017
Accepted post on December 26th, 2017

or updates an existing web page with user supplied data using a browser Application programming interface (API) that can create JavaScript. XSS allows attackers to execute scripts in the victim's browser which can hijack user sessions, deface web sites, or redirect the user to malicious sites.

4. Broken Access Control: Restrictions on what authenticated users are allowed to do are not properly enforced. Attackers can exploit these flaws to access unauthorized functionality and/or data, such as access other users' accounts, view sensitive files, modify other users' data, change access rights, etc.

5. Security Misconfiguration: Good security requires having a secure configuration defined and deployed for the application, frameworks, application server, web server, database server, platform and so on. Secure settings should be defined, implemented, and maintained, as defaults are often insecure. Additionally, software should be kept up to date.

6. Sensitive Data Exposure: Many web applications and APIs do not properly protect sensitive data, such as financial, healthcare, and Personally identifiable information (PII). Attackers may steal or modify such weakly protected data to conduct credit card fraud, identity theft, or other crimes. Sensitive data deserves extra protection such as encryption at rest or in transit, as well as special precautions when exchanged with the browser.

7. Insufficient Attack Protection: The majority of applications and APIs lack the basic ability to detect, prevent, and respond to both manual and automated attacks. Attack protection goes far beyond basic input validation and involves automatically detecting, logging, responding, and even blocking exploit attempts. Application owners also need to be able to deploy patches quickly to protect against attacks.

8. Cross-Site Request Forgery (CSRF): A CSRF attack forces a logged-on victim's browser to send a forged HTTP request, including the victim's session cookie and any other automatically included authentication information, to a vulnerable web application. Such an attack allows the attacker to force a victim's browser to generate requests the vulnerable application thinks are legitimate requests from the victim.

9. Using Components with Known Vulnerabilities: Components, such as libraries, frameworks, and other software modules, run with the same privileges as the application. If a vulnerable component is exploited, such an attack can facilitate serious data loss or server takeover. Applications and APIs using

components with known vulnerabilities may undermine application defenses and enable various attacks and impacts.

10. Underprotected APIs: Modern applications often involve rich client applications and APIs, such as JavaScript in the browser and mobile apps, that connect to an API of some kind (SOAP/XML, REST/JSON, RPC, GWT, etc.). These APIs are often unprotected and contain numerous vulnerabilities. Our proposed solution to detect and prevent those above hacking techniques is using Decision Tree algorithm to apply data mining into Web Application Firewall. The main idea is that based on the input package dataset with class labels 1 - anomalous and 0 - normal (we call this the training phase), we built a tree model which can predict whether a new package is normal or anomalous.

Decision tree is a type of supervised learning algorithm that is mostly used in classification problems, which supervision comes from the class labels in the training data set in learning phase. A decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute value, each branch represents an outcome of the test, and tree leaves holds a class label. [3]

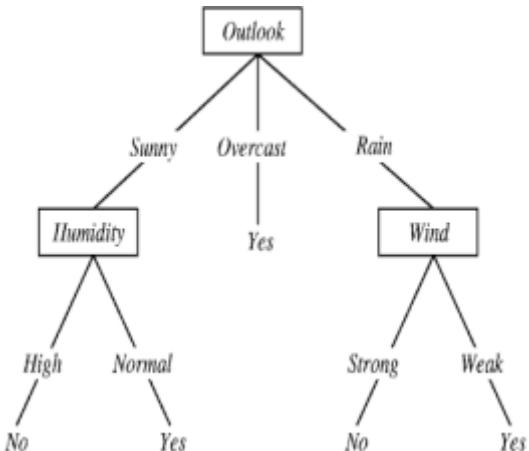


Figure 1. Decision tree model of Play golf dataset

III. PLAN IMPLEMENTATION

1. Normalize data and Select features

We used HTTP Dataset CSIC 2010 developed at the "Information Security Institute" of CSIC (Spanish Research National Council), which is generated automatically and contains 36,000 normal requests and more than 25,000 anomalous requests. The HTTP requests are labeled as normal or anomalous and the dataset includes attacks such as SQL injection, buffer overflow, information gathering, files disclosure, CRLF injection, XSS, server side include, parameter tampering and so on [4]. Each HTTP request in the dataset contain the following features: method, URL, protocol,

userAgent, pragma, cacheControl, accept, acceptEncoding, acceptCharset, acceptLanguage, host, connection, contentLength, contentType, cookie and payload. Especially, HTTP POST requests have other features: content-length and payload.

```
GET http://localhost:8080/tienda/publico/anadir.jsp?id=3&nombre=Vino+Rioja&precio=100&cantidad=55&Bt=A#Fadir+al+carrito
HTTP/1.1
User-Agent: Mozilla/5.0 (compatible; Konqueror/3.5; Linux) KHTML/3.5.8 (like Gecko)
Pragma: no-cache
Cache-control: no-cache
Accept: text/xml,application/xml,application/xhtml+xml,text/html;q=0.9,image/png,*/*;q=0.5
Accept-Encoding: x-gzip,x-deflate,gzip,deflate
Accept-Charset: utf-8,utf-8;q=0.5,*;q=0.5
Accept-Language: en
Host: localhost:8080
Cookie: JSESSIONID=B7161ACA043B0E6014CA42A4BCD06AB5
Connection: close
```

Figure 2. Normal GET traffic package from CSIC 2010 Dataset

```
GET https://localhost:8080/tienda/publico/anadir.jsp?id=2&nombre=Jam%F3n+ib%F3rlico&precio=85&cantidad=%27%3B-DROP+TABLE
+usuarioss3B+SELECT+*+FROM+datos+WHERE+nombre+LIKE+'%27%3B+At%F3adir+al+carrito+HTTP/1.1
User-Agent: Mozilla/5.0 (compatible; Konqueror/3.5; Linux) KHTML/3.5.8 (like Gecko)
Pragma: no-cache
Cache-control: no-cache
Accept: text/xml,application/xml,application/xhtml+xml;text/html;q=0.9,image/png,*/*;q=0.5
Accept-Encoding: x-gzip,x-deflate,gzip,deflate
Accept-Charset: utf-8,utf-8;q=0.5,*;q=0.5
Accept-Language: en
Host: localhost:8080
Cookie: JSESSIONID=B92A8B48B9008CD19F622A994E0F650
Connection: close
```

Figure 3. Anomalous GET Traffic Package (SQL Injection Attack) from CSIC 2010 Dataset

According to the GeFSCFS measure (The generic feature selection measure with the correlationfeature-selection measure proposed by Hai Thanh Nguyen [5]), we normalize and select 11 features from the CSIC 2010 dataset as shown in Figure 4. These 11 features are: Length of the arguments, Number of arguments, Number of digits in the arguments, Number of other char in the arguments, Number of digits in the path, Number of special char in the arguments, Length of the path, Length of the request, Number of letters in the arguments, Number of letters char in the path and Number of special char in the path.

Feature Name	Feature Name
Length of the header "Accept-Charset"	Length of the path *
Length of the arguments *	Length of the header "Accept"
Length of the header "Accept-Encoding"	Length of the request *
Length of the header "Accept-Language"	Length of the header "Cookie"
Length of the header "Content-Length"	Length of the header "Content-Type"
Length of the Host	Length of the header "Referer"
Length of the header "User-Agent"	Method identifier
Number of arguments *	Number of letters in the arguments *
Number of digits in the arguments *	Number of distinct bytes
Number of other char in the arguments *	Number of letters char in the path *
Number of digits in the path *	Number of 'special' char in the path *
Number of other char in path	Number of cookies
Maximum byte value in the request	Minimum byte value in the request
Number of 'special' char in the arguments *	Entropy
Number of keywords in the path	Number of keywords in arguments

Figure 4. Features from experts knowledge for Web attack detection; * is 11 features selected by GeFSCFS measure

To verify the feature selection, we experience two circumstances to see the difference between the data distribution of selected and unselected feature and the error rate if choose that feature for prediction. The anomalous package be examined is in the Figure 3.

- With Method identifier feauture: there are 43,088 GET requests/54508 requests that normal request is 28.000 (65%) and anomalous is 15,088 (35%) → Examined request was predicted as normal while it is anomalous. As the result, the error rate was 65%.
- With Number of digits in arguments (in the examined request is 15 digits) feature: there are 158 requests that has the same number of digits in arguments/54508 requests, and all 158 requests are anomalous. As the result, the error rate was 0%.



Figure 5. Distribution of Number of Digits in Arguments feature in CSIC 2010 Dataset

2. Build tree

When building the tree model, there are two problems need to be solved: what is the best split for the data and when should the splitting stop. Choosing the best split:

$$\text{Gini}(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2$$

Where:

- t: is nodes
 - c: is the number of classes
 - p(i|t): the fraction of records belonging to class i at a given node t.
- Decision Tree stop condition

A stopping condition is to terminate the tree-growing process. In this project, we will continue expanding a node until it met one of these conditions: all the records belong to the same class, number of failure at the node is too small or node reaches max depth of the tree.

Calculate Gini Impurity Process

Observing that all of our 11 selected features are continuous variables, so for each node of a tree, we performed a Gini impurity calculating to decide the best split of tree until met the stop condition. The process includes these steps:

- Step 1: Sort distinct value of the feature into an array
- Step 2: Calculate the average of two consecutive elements in the array
- Step 3: For each average calculated value, divide array into two small sets then use the formula given

above to calculate Gini impurity of each set.

- Step 4: To calculate the whole Gini impurity of feature, multiple each Gini impurity to percent of its set and sum two final result.
- Step 5: Compare Gini impurity corresponding to the average value to find the smallest value. This Gini and average value will be the best split for this feature.

Repeat all five steps to calculate the Gini impurity of all features. Then choose feature with smallest Gini impurity for the node with best split value we have calculated. After this, we will have a node, which will divide data set to two smaller sets base on split value. The same process is applied for the two smaller sets until the stop condition is met. It also means that all of smaller data set you divided has been a leaf.

IV. ANALYSIS

After selecting the features, we use the HTTP Dataset CSIC 2010 to build tree and test our algorithm. Using rpart function in R package's rpart, we have constructed a tree, which structure is shown in Figure 7.

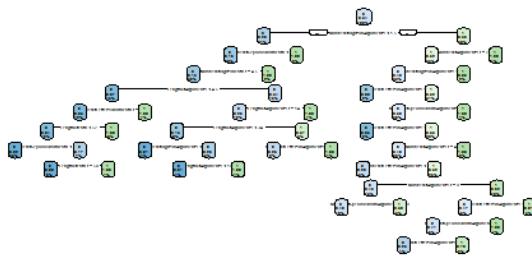


Figure 7. Our Decision tree of the HTTP Dataset CSIC 2010

To test the efficient of our tree model, we used the holdout method, which means divided the dataset into two subsets: 2/3 for training phase and 1/3 for testing phase.

The accuracy of the result is upto 93.5%, which we beleive that the algorithm is qualified enough for our goal. Detailed results are shown in Figure 8 and Table 1.

```
> predict_data<- predict(rtree_dictionary$`6`, test_data, type = "class")
> sum(predict_data==test_data$isAttack)/length(predict_data)
[1] 0.9353573
```

Figure 8. Decision Tree Accuracy Result

Table 1. A Confusion Matrix Of Model

		Actual	
		Normal	Anomalous
Predict	Normal	34722	1473
	Anomalous	1278	5084

From Table 1, we can calculate that the sensitivity of model is 96.45% and the specificity is 77.54%.

V. EXPERIMENTAL RESULTS

Our WAF model includes three separated parts: HTTP Module, Admin App and R Server. The HTTP Module will be deployed in the target website's web sever, while Admin App and R Server will be deployed in a system call Admin system. This model allows us to become third party that provides security services, which will effect to customer's web server as least at possible. However, it poses some challenges for our system such as:

- Network traffic between the module and Admin system
- Security problems about the connection between module and Admin system
- The pressure on the system affordable with the number of modules

Workflow of this module

Include five steps:

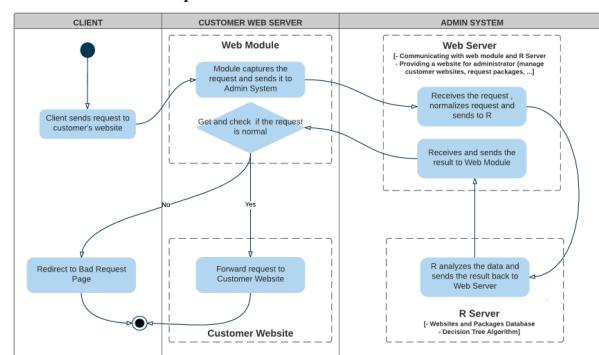


Figure 9. Workflow of the WAF

- Step 1: The module captures the users request before it is sent to target website.
- Step 2: Module sends captured request to Admin System through Web Api.
- Step 3: Inside the Admin System:
 - Web Server receives, normalizes and sends the request to R Server.
 - R Server analyzes the received data based on Decision Tree algorithm and returns the result to Web Server.
 - Web Server receives analyzed result.
- Step 4: The Admin System sends result to the module.
- Step 5: Base on the received result, the module will decide how to process the request package.

Result

To verify it works, we deployed the module on the website that existed sql injection risk.

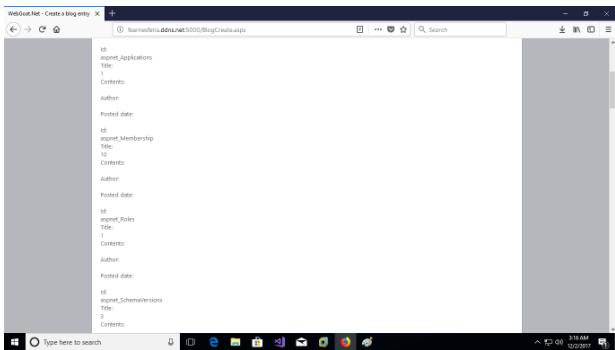


Figure 10. The hacker has performed injection to this site “BlogCreate.aspx” to get list table of database

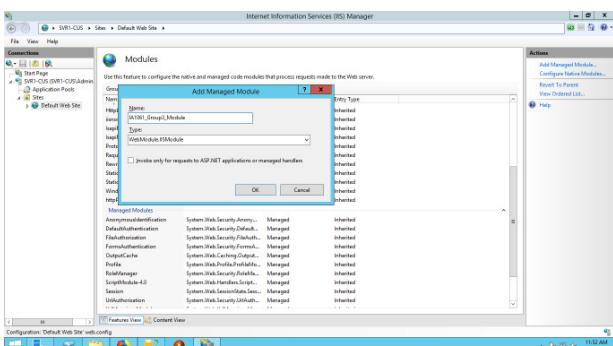


Figure 11. Add module to customer IIS Server

After collecting more than 4000 packages of this site (most of them are packages send to the site BlogCreate.aspx), we try to inject database again with the same query. The returned result is Bad Request.



Figure 12. Result after deploy module

VI. CONCLUSION

With the result received in above section, we believed that our model is working well. The WAF has satisfied our identified problem: determine whether a package is normal or anomalous based on its header. Moreover, with the explosive growth of data, we think it is a good and potential idea to apply data mining to security field.

ACKNOWLEDGEMENT

We would like to thank Mr. Duong Thien Tu and

Mr. Nguyen Sieu Dang for helping us complete this project.

REFERENCES

- [1] Jnena, R. M. (2013). Modern Approach for WEB Applications Vulnerability Analysis (Doctoral dissertation, The Islamic University of Gaza).
- [2] (2017) OWASP ConOWASP Top 10 - 2017 The Ten Most Critical Web Application Security Risks. Available at: https://www.owasp.org/images/7/72/OWASP_Top_10-2017_%28en%29.pdf.pdf
- [3] Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.
- [4] HTTP Dataset CSIC 2010. Available at: <http://www.isi.csic.es/dataset/>
- [5] Nguyen, H. T. (2012). Reliable Machine Learning Al-gorithms for Intrusion Detection Systems. Faculty of Computer Science and Media Technology.
- [6] Tan, P. N., Steinbach, M., & Kumar, V. (2006). Classification: basic concepts, decision trees, and model evaluation. Introduction to data mining, 1, 145-205.

Intent Analyzer, Dialog flow Controller and Sentence Suggestion: *Location-based Chat bot team*

Tuong Duc Khang
FPT University
HCMC, Viet Nam

Nguyen Van Hung
FPT University
HCMC, Viet Nam

Duong Hai Quang
FPT University
HCMC, Viet Nam

Thai Hoang Duy Bao
FPT University
HCMC, Viet Nam

Abstract

The problem is given the bot needs to understand what user texted, extracts correct information and response correctly. There are currently many of APIs on the Internet can help recognize a sentence, but they are not fit in this project due to having no dialog flow control. Since the bot needs to manage the conversations through dialogs, a connection between intent and dialog need to be established. Therefore, a homemade intent analyzer and a dialog flow controller are built locally, on the same machine. This can reduce the time spend on requesting to a third-party API and receiving returned information. We can also customize how exactly the bot should analyze, add more functions to help it recognize better. In other words, to make it more intelligent.

Keywords

Intent analyzer; Dialog flow controller

I. INTRODUCTION (HEADING 1)

The problem of analyzing a human sentence has been raising from the beginning of AI era. However, it is a sub problem in a bigger problem: how to teach robots to talk or at least response by text like a human. But why do we need to build a human-like robot? The answer is “Humanoid robots result in good science. They push the boundaries of biology, cognitive science, and engineering, generating a mountain of scientific publications in many fields related to humanoid robotics, including: Computational neuroscience, A.I., speech recognition, compliant grasping and manipulation, cognitive robotics, robotic navigation, perception, and the integration of these amazing technologies within total humanoids.” [1]

There have been several teams studying about this problem and they have been introducing several solutions.

Some solutions are implemented as a web API to help everyone to use them, such as <https://dialogflow.com/> and <https://www.luis.ai/> (from Microsoft). These services can help us enough to recognize and extract information from an input sentence. But is it enough?

As we can see (and feel), humans talking to humans is not as simple as make a request and receives a response. A conversation between two people can be numerous sentences combining. Each sentence can make the conversation turn to a different way. That can lead to another problem: a sentence doesn't have the same meaning at all time. Let's have an example. Let's take a simple example. A female customer walks in the fashion shop. She has some conversations with the shop employees. A man passes by, not looking at the woman and he hears the woman says: “I don't like it”. The question is can the man understand what thing that the woman does not want? Of course, the answer is no. She didn't like the coat, or she didn't like the shirt? Without context, that man can't give exact information about that sentence. However, the shop employee may be different. Because that employee has been talking to that woman and recommending her a shirt, the employee can understand that the woman doesn't like the shirt. That is the basic difference between not in a context and being in a context. So, what does it mean to have an intent analyzer and dialog flow controller coming together?

The intent analyzer will be responsible for detecting user's sentences, extract information from them and return information to the dialog flow controller. The dialog flow controller will be responsible for controlling dialogs. Each dialog has a session that save the current context and information to have it proceed correctly. If user change topic, a new dialog is added to the dialogs stack and the bot will start the new dialog, until it is end and return that previous dialog. A standalone intent analyzer can't handle a

Supervisor: Lam Huu Khanh Phuong, M.E
(Lecturer of IT Department, HCMC FPT University)
Defended topic in front of the scientific council of HCMC
FPT University on December 14th, 2017
Accepted post on December 26th, 2017

conversation in a higher level, where information of each sentences need to be stored for further uses. Because of that, two components need to be connected.

There have been a lot of introduction about the sentence analysis. The most basic one is to use conditional method to check whether the input sentence is found in the studied sentences corpus, which means we can call if...else multiple times in the program, until they can cover basic sentences. This method is the very basic method and very simple to implement. However, this is also the most time-consuming and hopeless method since the number of sentences is large enough to make a developer tired of typing and thinking about new sentence.

A better solution is find a phrase in a sentence which the bot has learned. The better part is that we don't have to type so many similar sentences. We can reduce the work by finding the common phrases between those sentences, grouping them and then tell the bot that if the input sentence includes those phrase, mark it as recognized. This solution may sound to be a very good approach to solve the problem, or doesn't it?

The listed two approaches are very bad, from idea stage to implementation. The basic thing that make those solutions bad is the bot is not learning at all. When people change the way of using the language, we must programmatically update the bot to help it understand the modern language. Furthermore, the data (bot internal sentence corpus) are embedded straight in the code so what happen if the coding language we use today won't exist in the future. The whole data need to be re-create again and that will slow the bot learning process down.

In this paper, we concentrate on studying how to detect and extract information from a sentence and how to control the dialog sequence to make it natural. In section 2 we formalize the problem and describe our approach to solve it. Section 3 describes the data structures that need to be set up and the subsequent algorithms. We analyze our algorithms in Section 4. Experimental results and the software used are described in the concluding Section 5.

II. PROBLEM AND SOLUTION PLAN

Given a sentence "I would like 2 hamburgers", the bot must analyze this sentence and make clear what user wants to do. To do that, there are 3 problems:

Problem 1: How to know that it is an order request. A normal request order sentence contains some common phrase like "I would like" or "I want". So, if we declare these phrase in the bot and tell it if a sentence contains "I would like" or "I want" then it must be an order request. However, this solution

has a great weak point. If user didn't say "I would like" but saying "My mother would like" instead, we have to create one more phrase (which has the same intent – to order). We will fix this point in the next problem.

Temporary solution 1: create common phrase for bot to understand.

Problem 2: How to distinguish words from a sentence to identify which words belong to quantity and which word belong to product name

"2 hamburgers" doesn't come always. Other request can be "2 coke", "1 sandwich" or even "three sandwiches, please". With the previous solution, we will need to add more three phrases for bot to recognize. Therefore, that solution is not possible. Instead, we can generate separate group of words that have the same meaning. For example, "I/My mother/My uncle/He" should be in a group of words called pronoun. And these groups are called Entity. Next, a sentence won't be analyzed by phrases anymore. It will be analyzed by having a sequence of ordered Entity or not. In the previous example, to recognize "I would like 2 hamburgers", we will create a sentence that contains the following sequence of Entities:

[Pronoun] – [Request verb] – [Number] – [Words]
The reason of not creating "hamburger" in the "Product name" Entity or something like that is because we can't predict what product user wants to order. It can be "salad", "cheeseburger" or "coke". If we create an Entity including product names, it will be a very long list, not to mention that new food is created every day. For that reason, we will define product name as an unknown phrase.

Temporary solution 2: we will create an object called Pattern. A pattern is a sequence of ordered Entities and it is used to analyze user's input. All the information is stored in the database (storing in code just make it worse). This solution still has a weak point. See problem 5.

Problem 3: Manage dialog flow

We have already had tools to analyze a sentence now. But as the Introduction section represent, human's conversations include lots of series of sentences and we need to find a way to store information from the conversation. If the bot recommends a shirt to a customer and ask whether the customer like it or not, the bot must remember that it is asking the customer about the shirt and the next sentence of the customer is about that shirt. Therefore, if a customer says "I don't like it", the bot must know that the customer didn't like the shirt it just recommended. Without dialog flow control, the bot will end up "What don't you like?"

The basic idea of dialog flow control is simple: when a user says something and if the bot detected that sentence belongs that which conversation, it will start the conversation. Before that, the state of current conversation is stored, and the conversation is paused until the new conversation end.

Temporary solution 3: we will call a conversation is a Dialog from now on. There will be a dialog stack, which stores the current dialogs of user, to make sure all information are stored without being removed due to dialog changing. If user suddenly change to the previous dialog, the new dialog is removed, and the previous dialog continues. This solution can be capable for solving the main problem of chat bot. But there is a point we need to improve.

Problem 4: A lot of sentences have the same meaning “I want hamburger”, “I would like hamburger, please” and more sentences have the same meaning that the user wants to order hamburger. The problem with solution 2 is it doesn’t group the patterns those have the same meaning. This is a technical problem rather than a natural problem. Since we are going to use relational database, if we let one pattern trigger one dialog then there would be redundancy in database storing. Therefore, instead of letting pattern trigger dialog, we want to introduce a final new concept: Intent. An intent is a group of patterns those have the same intention and it will be responsible for triggering the dialog.

Temporary solution 4: Divide patterns into groups called intent. The analysis should return the information of the intent instead of pattern, so the bot can understand what user wanted.

Overall solution: Adopt concept about entity, pattern, intent and dialog. There will be two components: Intent analyzer to analyze user’s input and return the corresponding intent; dialog flow controller to manage the current dialogs. There will be an extra function using Levenshtein distance to suggest the sentence that it can understand when user texts something weird.

III. PLAN IMPLEMENTATION

We will create a database like Figure 1. Example data is shown in Figure 1.1

A pattern will have a lot of entities and vice versa. A pattern also can be checked that it needs to be matched from the beginning and at the end of the input sentence or not. One pattern has only one group and this “group” definition will be explained in the next section. Notice that one pattern can belong to one intent only. This means one sentence can only have one intention. It doesn’t conflict with the problem introduced above because one sentence can have several meanings but the intention it give is

only one. We are not including figurative meanings here so one sentence can have one intention only. An intention then can trigger one dialog only or not trigger anything. Multiple dialogs into one intents can cause conflict in the bot itself because it can’t decide which dialog should it trigger, especially in neutral context, which gives no clue for dialog prediction. Next, we will describe how intent analyzer works.

Given a sentence “I would like 2 hamburgers”, the goal of the chat bot is to return three things: the intention of user is to order; the product user wanted to order is hamburger; the quantity of hamburger is 2. The intent analyzer will handle this task and then the dialog flow controller will add the order dialog, a dialog used for handling order request, to the dialog stack and start it. After it is started, the information is stored and the bot response “Okay, 2 hamburgers”. After that the user says “Finish”, the chat bot calculates the prices and return the total price for this order. In figure 2, we have a diagram to show every step the bot takes to analyze the sentence. The first four steps are simply database query and sorting data. Below is detail description from step fifth. To check whether entity is found in the user’s input or not, we use regular expression. The limitation of string comparer is it can’t recognize numbers and words. Numbers can vary from 1 to 100 or 10000 and words can be any meaning words, phrase in language. If we use string comparer, we will need to define lots of number and lots of words, which are nearly infinitive. Therefore, we use \d+ and .*? in regular expression instead. There are four expressions we will use in this case:

- (1) (?:^|\W)^{expression_here}\$(?:^|\W)
- (2) (?:^|\W){expression_here}\$(?:^|\W)
- (3) (?:^|\W)^{expression_here}(?:^|\W)
- (4) (?:^|\W){expression_here}(?:^|\W)

The first expression is used for pattern which needs to be match from the beginning and at the end of input. The second pattern is used for patterns requiring matching at the end of input. The third expression is used for those requiring from the beginning and the last one is used for those doesn’t require anything.

Assume that the bot has looped to this pattern and this pattern doesn’t require match from the beginning or at the end:

[Pronoun] – [Request verb] – [Number] – [Words]
For the first entity, the analyzer will get all the pronouns. The words for pronoun in database are stored in this format:

I|He|She

Each word or phrase are separated by a slash (|). A slash in regular expression is like an “or” keyword

in coding. When loaded, the bot will check the input sentence for this regular expression:

(?:^|\W)I|He|She(?::\$|\W)

This expression means “find in the sentence whether it is started by one of the three words “I”, “He” or “She”. In the example sentence, result will be True and it return the words that match the expression, given one information: the pronoun in the sentence is “I”. The next entities are examined the same way. The result is shown in figure 3. Then the bot compares the quantity of matched entities in this pattern. If this quantity is larger than the quantity matched entities of previous pattern, then the bot will take this pattern as the final result.

But analyzer won’t detect and understand all sentences user has input. Therefore, there must be a sub function that help return the closest sentence that the bot can understand. In that case, a Levenshtein distance approach is used.

Given 2 strings A[1, 2, ..., n] and B[1, 2, ..., m] have the length are n & m. Levenshtein distance is a string metric for measuring the different between two strings. Informally, the Levenshtein distance between two strings is the minimum number of single-character edits (insertion, deletion and substitutions) required to change one word into the other.

$$\text{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \text{lev}_{a,b}(i-1,j) + 1 \\ \text{lev}_{a,b}(i,j-1) + 1 \\ \text{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

Example: Levenshtein distance between “Saturday” and “Sunday” is 3. For detail description, see Figure 4.

Saturday → Sturday → Surday → Sunday

For the dialog flow controller, each Dialog will have three important elements:

Status: “new”, “paused” or “ended”

Step: the current step of dialog

Exception: the current special step of dialog

The status can be “new”, “paused” or “end”. When the chat bot starts a new dialog, that dialog will be pushed to a dialog stack and the status is set to “new”. When it start another dialog, the current dialog is set to “paused”, step back one step and the new dialog is pushed to dialog stack. When the new dialog is ended, it will be pop out the stack and the previous dialog continues from the saved step.

The reason why dialog needs to step back one step when it’s paused is because it’s organized by “one step receive and one step response”, which means the first step is always a receiving message step; the next step is response to user and so on. When it receives a text, it will proceeds business logic and continues the response step. After responding it will

increase the step to the next receiving step and wait for user’s input. Therefore if a dialog is paused, it needs to step back one step, to the previous response step in order to repeat the response to user when the dialog is resumed.

Exceptions are special steps that can go through several steps at one time. For example, in a normal conversation, when user wants to order, the bot will ask for the product → user says the name of product → the bot asks for quantity → user response quantity. An exception occurs when user immediately tells the bot the product name and quantity at the same time (“I would like 2 hamburgers”). In this case the bot needs to apply different logic to handle two information. After finishing an exception, the dialog will continue to the next normal step, which is defined in the exception function.

A special case is that user wants to start a conversation which is paused before and pause the current conversation. In this case, the bot will pause the current dialog, save its status, get the previous dialog out of stack and push it to front again. Detail description in figure 5.

IV. ANALYSIS

This method works well on separating words. However this doesn’t apply business logic, such as converting number in full string type (“one”, “two”, and “three”...). Therefore, when a new pattern such as “I would like two hamburgers” is added, the intent analyzer will return the quantity is “two”, not “2”. Business logic is still implemented straight in code. That’s why there is a “group” definition. Patterns with different groups will be applied different business logic. The developer needs to create another function to do this task.

Next, the Levenshtein distance approach can help suggesting sentences but not always right. Since the sentence it suggests is loaded from database, which is not includes grammar checking. For example, when user texts “I wouldn’t like 2 hamburgers”, the bot will suggest the sentence “I would like 2 hamburgers” since those two sentences are very similar. However the meanings are completely different so the suggestion system doesn’t work very accurate.

V. EXPERIMENTAL RESULTS AND CONCLUSION

All the data is included in the project. User can follow the project start up instruction in the complete document. User can test the intent analyzer on the ChatbotManagement system, which is included in this project. For the dialog flow controller and sentence suggestion, user can chat with the Passio

bot.

The result of sentence suggestion is quietly inaccurate at first, since there is no data mining and analysis included. However, this is a potential function needs to be implemented as soon as possible to guide user to chat with chat bot. In the future this rate will be lowered when some mathematic statistic and probability techniques are applied. This also can be a method to attract user since it will never end chatting despite whatever user said.

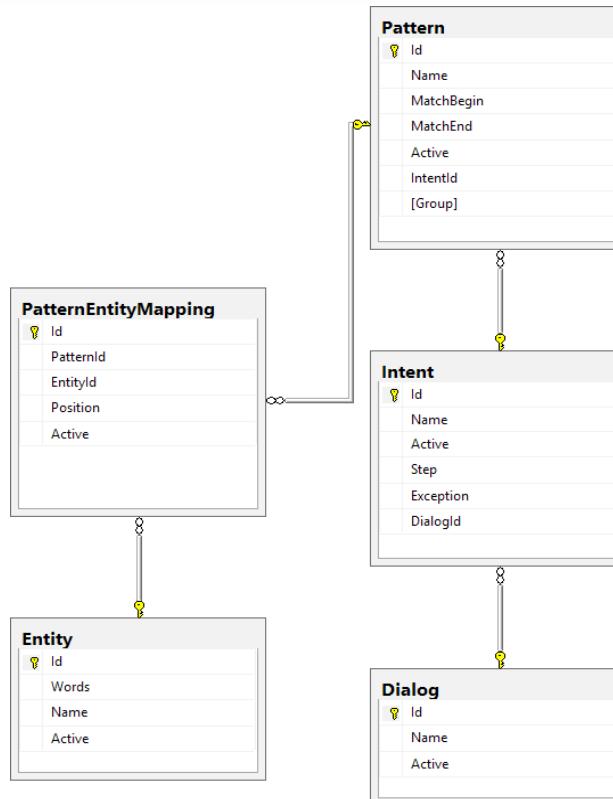
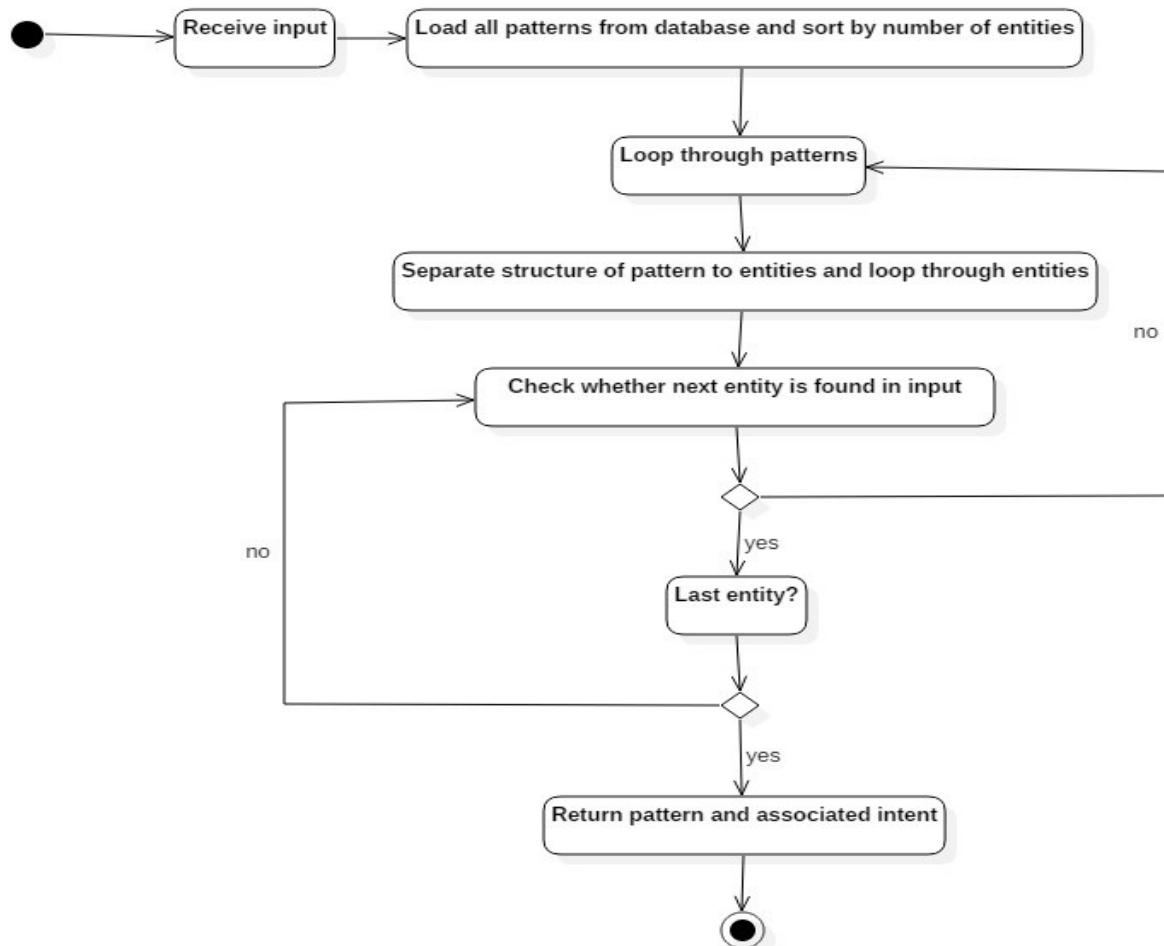
We conclude that initial results are encouraging our proposed solution and this approach can be used in practice.

AKNOWLEDGEMENT (HEADING 5)

We thank Mr.Lam Huu Khanh Phuong for help with the software. We thank the referees of this and the earlier conference version for several suggestions that helped improve the presentation significantly

REFERENCES

- [1] “Why shoule we build humanlike robots” by David Hanson.
<https://spectrum.ieee.org/automaton/robotics/humanoids/why-we-should-build-humanlike-robots>
- [2] Inspiration for designing the analyzer: api.ai (now changed to dialogflow)
<https://dialogflow.com/>
- [3] Inspiration for designing the dialog flow controller system: Microsoft bot framework
<https://dev.botframework.com/>
- [4] Levenshtein distance algorithm
https://en.wikipedia.org/wiki/Levenshtein_distance

**Figure 1. The implemented database****Figure 2. Intent analyzer process**

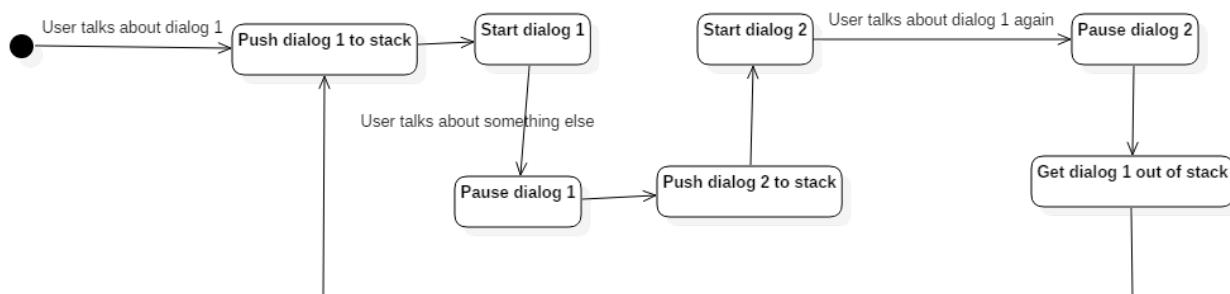
I would like 2 hamburgers|

Phân tích

```
{
  "IntentId": 6,
  "Group": 1,
  "Matches": {
    "Pronoun": "I",
    "Request verb": "would like",
    "Number": "2",
    "Phrase": "hamburger",
  }
}
```

Figure 3. Result after analyze sentence

		S	a	t	u	r	d	a	y
	0	1	2	3	4	5	6	7	8
S	1	0	1	2	3	4	5	6	7
u	2	1	1	2	2	3	4	5	6
n	3	2	2	2	3	3	4	5	6
d	4	3	3	3	3	4	3	4	5
a	5	4	3	4	4	4	4	3	4
y	6	5	4	4	5	5	5	4	3

Figure 4. Levenshtein distance algorithm**Figure 5.** Dialog flow for re-calling old dialog

sProcurement:

TOPSIS Method for Supplier Selection Problem

Hoang Vinh Son
FPT University
HCMC, Viet Nam

Nguyen Dinh Thong
FPT University
HCMC, Viet Nam

Vu Thai Nam
FPT University
HCMC, Viet Nam

Huynh Ngoc Dat
FPT University
HCMC, Viet Nam

Abstract

Procurement is the process of finding, evaluating, selecting, agreeing terms and acquiring goods, services from suppliers, often via price quotation. In today's highly competitive environment, an effective supplier selection process is very important to the success of any manufacturing organization. In this context, supplier selection represents one of the most important functions to be performed by the purchasing department. Supplier selection is a multi-criterion problem which includes both qualitative and quantitative factors (criteria). A trade-off between these tangible and intangible factors is essential in selecting the best supplier. A number of models and techniques have been developed to deal with selecting and evaluating suppliers. The aim of this study is developing a methodology to evaluate suppliers in supply chain cycle based on Technique for Order Preference by Similarity to Ideal Solution method (TOPSIS).

Keywords

*TOPSIS; Multi Criteria Decision Making;
Supplier Selection*

I. INTRODUCTION

In order to maintain a competitive position in the global market, organizations have to follow strategies to achieve shorter lead times, reduced costs and higher quality. Therefore, suppliers play a key role in achieving corporate competitiveness, and as a result of this, selecting the right suppliers is a critical component of these new strategies. Several conflicting quantitative and qualitative factors or criteria like cost, quality, delivery etc. affect supplier selection problem; therefore, it is a Multi-Attribute Decision Making (MADM) problem.

Several methods, such as the analytic hierarchy process (AHP), the analytic network process (ANP), the linear weighting methods (LW), total cost approach (TCA) and mathematical programming (MP) techniques, have helped decision makers (DMs) to deal with supplier selection problem. However, the problem of selecting suppliers is that most of the input information is not known precisely. That is why, the problem becomes more difficult and complicated. In these cases, the fuzzy sets theory is usually used for dealing with uncertainty. However, the disadvantage of fuzzy sets theory is that it cannot handle incomplete data and information [1]. To overcome the problem, Deng in 1989 [2] proposed Grey theory being an effective mathematical tool to deal with systems analysis characterized by imprecise and incomplete information [1]. So, the advantage of grey theory over fuzzy theory [3,4] is that grey theory, in addition to the condition of the fuzziness, can take incomplete data and information into consideration [11,12]. Grey theory is based on the degree of information, known [1,2]: if the system information is unknown, it is called a black system; if the information is fully known, it is called a white system; and eventually, if the information is known partially, it is called a grey system.

In grey theory, there are two famous methods proposed by Deng [2] and Li et al. [1]. The two methods are used for MADM problems; evaluating and ranking alternatives against some factors or criteria. The methods are similar to TOPSIS method. Deng's [2] method is based on the minimization of maximum distance from the ideal referential alternative. Zhang et al. [5] presented the method of Deng [2] as a means to reflect uncertainty in multiple attribute models through interval fuzzy numbers.

Li's et al. [1] grey based approach, applied for supplier selection problem, calculates a grey possibility degree between compared suppliers' alternatives set and ideal referential supplier alternative to determine the ranking order of all alternatives of supplier and to select the ideal supplier based on grey numbers. Although the two methods [11,12] are similar to TOPSIS, they neglect the negative ideal referential alternative to evaluate and rank the alternatives.

Supervisor: Lai Duc Hung, M.E
(Lecturer of IT Department, HCMC FPT University)
Defended topic in front of the scientific council of HCMC
FPT University on December 14th, 2017
Accepted post on December 26th, 2017

Based on TOPSIS or the concept of TOPSIS the chosen alternative should have the shortest distance from the positive ideal solution (PIS) and the farthest from the negative ideal solution (NIS) [6]. For tackling the problem of Li's et al and Deng's, Jadidi et al. [7] proposed a new technique by using TOPSIS concept. In fact, the authors improved the methods of Li et al. [1] and Deng [2] by this concept. In the article, Jadidi et al. [7] compared his method with the methods of Li et al. and Deng and demonstrated that the new method based on TOPSIS concept has a more optimal solution.

In this article, instead of using TOPSIS concept, we use TOPSIS method itself to evaluate and select the best supplier in grey theory. The relative closeness between each of the alternative sequence and the referential sequences (positive and negative ideal solutions) is calculated to compare the ranking of grey numbers and select the most desirable supplier.

II. PROBLEM AND SOLUTION PLAN

By TOPSIS, developed by Hwang and Yoon [6] for evaluating and ranking alternatives against some factors, the chosen alternative has the shortest distance from PIS and the farthest from NIS.

In order to apply TOPSIS for evaluating suppliers we use Li's et al. [1] article assumption, data, information, and notation (see [11, 18]).

Step 1: Normalize the decision matrix.

Create an evaluation matrix consisting of m alternatives and n criteria $\Rightarrow (X_{ij})_{m \times n}$

We introduce the following formulas to normalize each attribute value x_{ij} in decision matrix $X = (x_{ij})_{m \times n}$ into a corresponding element r_{ij} in normalized decision matrix:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m (x_{ij})^2}} ; i = 1, 2, \dots, m ; j = 1, 2, \dots, n$$

Step 2: Calculate the weighted normalized decision matrix. Matrix $(W_j)_n$ is the weight vector of the criteria.

$$i = 1, 2, \dots, m ; j = 1, 2, \dots, n$$

Step 3: Determine the positive and negative ideal solutions. The PIS A^+ and NIS A^- are determined, respectively, as follows:

$$t_{ij} = r_{ij} \cdot w_j ; i = 1, 2, \dots, m ; j = 1, 2, \dots, n$$

$$\text{and } A^- = \{t_1^-, t_2^-, \dots, t_n^-\}$$

$$\text{Where } t_j^+ = \max_{1 \leq i \leq m} \{t_{ij}\} (j \in n)$$

$$\text{and } t_j^- = \min_{1 \leq i \leq m} \{t_{ij}\} (j \in n)$$

Step 4: Measure the distance from positive and negative ideal solutions.

The separation of each alternative from the PIS, S_i^+ , is given as

$$S_i^+ = \left(\sum_{j=1}^n (t_{ij} - t_j^+)^2 \right)^{\frac{1}{2}}, i \in m$$

Similarly, the separation forms the NIS, S_i^- , is given as

$$S_i^- = \left(\sum_{j=1}^n (t_{ij} - t_j^-)^2 \right)^{\frac{1}{2}}, i \in m$$

Step 5: Calculate the closeness coefficient to the ideal solutions. The closeness coefficient of the I alternative A_i with respect to the ideal solutions is defined as

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-}, i \in M$$

Step 6: Rank the preference order.

A set of alternatives then can be ranked by preference according to the descending order of C_i ; in other words, larger C_i means better alternative.

III. PLAN IMPLEMENTATION

Suppose, there are 5 suppliers that are selected for evaluation: A_1, A_2, A_3, A_4, A_5 . There are 5 criterions: X_1, X_2, X_3, X_4, X_5 (Evaluation scores are given on 5-point scale).

Symbols	Meaning				
	X1	X2	X3	X4	X5
A1	2	4	3	1	4
A2	4	3	4	2	3
A3	4	3	2	4	4
A4	2	2	3	4	5
A5	4	2	3	2	3
Weight	0.2	0.2	0.1	0.4	0.1

Sets of data are collected from statistics

We have matrix $(X_{ij})_{m \times n}$. With n is the number of criterion, m is the number of suppliers.

Step 1: Normalize the decision matrix.

Example: $i = 1, j = 1$

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m (x_{ij})^2}} \Rightarrow r_{11} = \frac{x_{11}}{\sqrt{\sum_{i=1}^5 (x_{11})^2}} = \frac{2}{\sqrt{2^2 + 4^2 + 4^2 + 2^2 + 4^2}} = 0.0357$$

After doing the same calculation, we have matrix

(Rij)m x n

	X1	X2	X3	X4	X5
A1	0.035 7	0.095 2	0.063 8	0.024 4	0.053 3
A2	0.071 4	0.071 4	0.085 1	0.048 8	0.040 0
A3	0.071 4	0.071 4	0.042 6	0.097 6	0.053 3
A4	0.035 7	0.047 6	0.063 8	0.097 6	0.066 7
A5	0.071 4	0.047 6	0.063 8	0.048 8	0.040 0
Weight	0.2	0.2	0.1	0.4	0.1

Step 2: Calculate the weighted normalized decision matrix. Matrix (Wj)n is the weight vector of the criteria.

Example: i = 1, j = 1

$$t_{ij} = r_{ij} \cdot w_j \Rightarrow t_{11} = r_{11} \cdot w_1 = r_{11} \cdot 0.2 = 0.0071$$

After doing the same calculation, we have matrix (Tij) m x n:

	X1	X2	X3	X4	X5
A1	0.007 1	0.019 0	0.006 4	0.009 8	0.005 3
A2	0.014 3	0.014 3	0.008 5	0.019 5	0.004 0
A3	0.014 3	0.014 3	0.004 3	0.039 0	0.005 3
A4	0.007 1	0.009 5	0.006 4	0.039 0	0.006 7
A5	0.014 3	0.009 5	0.006 4	0.019 5	0.004 0
Weight	0.2	0.2	0.1	0.4	0.1

Step 3: Determine the positive and negative ideal solutions. The PIS A+ and NIS A- are determined, respectively, as follows:

Example: j = 1

$$t_j^+ = \max_{1 \leq i \leq m} \{t_{ij}\} \Rightarrow t_1^+ = \max_{1 \leq i \leq m} \{t_{i1}\} = 0.0143$$

$$t_j^- = \min_{1 \leq i \leq m} \{t_{ij}\} \Rightarrow t_1^- = \min_{1 \leq i \leq m} \{t_{i1}\} = 0.0071$$

After doing the same calculation, we have:

$$A^+ = \{0.0143, 0.0190, 0.0085, 0.0390, 0.0067\}$$

And

$$A^- = \{0.0071, 0.0095, 0.0043, 0.0098, 0.0040\}$$

Step 4: Measure the distance from positive and negative ideal solutions.

Example: i = 1

$$S_i^+ = \left(\sum_{j=1}^n (t_{ij} - t_j^+)^2 \right)^{\frac{1}{2}} \Rightarrow S_1^+ = \left(\sum_{j=1}^5 (t_{1j} - t_j^+)^2 \right)^{\frac{1}{2}} =$$

$$\sqrt{(0.0071 - 0.0143)^2 + (0.0190 - 0.0190)^2 + \dots + (0.0053 - 0.0067)^2} = 0.0302$$

$$S_i^- = \left(\sum_{j=1}^n (t_{ij} - t_j^-)^2 \right)^{\frac{1}{2}} \Rightarrow S_1^- = \left(\sum_{j=1}^5 (t_{1j} - t_j^-)^2 \right)^{\frac{1}{2}} =$$

$$\sqrt{(0.0071 - 0.0071)^2 + (0.0190 - 0.0095)^2 + \dots + (0.0053 - 0.0040)^2} = 0.0098$$

After doing the same calculation, we have:

$$S^+ = \{0.0302, 0.0203, 0.0065, 0.0121, 0.0022\}$$

$$S^- = \{0.0098, 0.0137, 0.0305, 0.0295, 0.0123\}$$

Step 5: Calculate the closeness coefficient to the ideal solutions.

Example: i = 1

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-} \Rightarrow C_1 = \frac{S_1^-}{S_1^+ + S_1^-} = \frac{0.0098}{0.0302 + 0.0098} = 0.2457$$

After doing the same calculation, we have:

$$C = \{0.2457, 0.4030, 0.8239, 0.7090, 0.0600\}$$

Step 6: Rank the preference order.

$$0.2457, 0.4030, 0.8239, 0.7090, 0.0600$$

$$\Leftrightarrow A1, A2, A3, A4, A5$$

Sort in descending order

$$0.8239 > 0.7090 > 0.0600 > 0.4030, > 0.2457$$

$$\Leftrightarrow A3 > A4 > A5 > A2 > A1$$

Supplier A3 is the best choice.

IV. EXPERIMENTAL RESULTS AND CONCLUSION

CONCLUSION

The study discussed that how to select the best supplier in supplier selection problems when decision makers set the target value of each criterion. Although many approaches can solve the problem, the study proposed a method and a procedure to extend the TOPSIS method to solve the problem. The main advantages of using TOPSIS method are “TOPSIS logic is rational and understandable”, “The computation processes are straightforward”, “The concept permits the pursuit of best alternatives criterion depicted in a simple mathematical” and “The importance weights are incorporated comparison procedures”. Due to this, decision making for selection of suitable supplier is of special importance. Acquired results from numerical example determine that this model could be used for decision making optimization in supplier selection.

ACKNOWLEDGMENT

This paper owes massive thanks to Mr. Lai Duc Hung at FPT University. He always supports us whenever we run into a trouble or have a question about our project. He consistently allowed this project to be

our own work, but steered us in the right direction whenever he thought we needed it.

REFERENCES

- [1] G.D. Li, D. Yamaguchi, M. Nagai, "A grey based decision making approach to the supplier selection problem," Mathematical and Computer Modelling 36 (2007), pp. 573–581.
- [2] J.L. Deng, "The introduction of grey system", The Journal of Grey System. 1 (1989), pp. 1–24.
- [3] L.A, Zadeh, "Fuzzy sets", Information and Control. 8 (1965), pp 338– 353.
- [4] R.E, Bellman, and L.A, Zadeh, "Decision making in a fuzzy environment", Management Science. 17 (1970), pp. 141–164.
- [5] J.J, Zhang, D.S, Wu, and D.L, Olson, "The method of grey related analysis to multiple attribute decision making problems with interval numbers", Mathematical and Computer Modelling. 42 (2005), pp. 991– 998.
- [6] C.L. Hwang and K. P.Yoon, Multiple attribute decision making methods and applications, Springer, New York (1981).
- [7] Jadidi O., Tang S. H., Firouzi F., Rosnah M. Y., "An optimal grey based approach based on TOPSIS concepts for supplier selection problem", International Journal of Management Science and Engineering Management. Vol. 4 (2009) No. 2, pp. 104-117.

University Admission Counseling System for High School Students

Ngo Nguyen Thuy Van

FPT University
HCMC, Viet Nam

Le Thanh Danh

FPT University
HCMC, Viet Nam

Vo Manh Hung

FPT University
HCMC, Viet Nam

Nguyen Le Minh

FPT University
HCMC, Viet Nam

Abstract

Each year, over 800,000 students are coming to university entrance exam, the need of information about universities are very necessary.

There are many websites provide admission information but the difficulty is Viet Nam has over 700 universities with too many information about scores, blocks, targets, and those website don't have the contracted data to finding....It makes students feel worry about their choices or too hard to finding information.

In additional, VietNam's Highschools don't have a good orientation for students. So that, students don't know which career is suitable for them.. The solution is build a system that helping students easy to find information about universities and consulting which career are suitable with their personality.

Keywords

Search university; MBTI test; recommendation; rating and review; report

I. INTRODUCTION

Main problem that the students nowaday can't find the fully and constructed information about all university to make a choice. Some university's information gates like kenhthuyensinh.vn, tuoitre.vn....All informations are put in article format so that it is difficult for students to search the informations that they want. The other university information systems face some problem like:

- Informations are too discursive.
- Don't have a clear information searching system
- Don't have a personality test.
- Don't have an answer and question function
- Don't have rating and review function

Students are really need informations about university's introduce, major, score and how other

peoples, specially is the students of that university think about it.

And the most importance, other information systems don't have personality test to help students know which career is suitable for them, give them an recommendation for suitable university.

Instead of a general information, we will focus more on searching, rating, review and recommendation by user's information. The system uses 3 criteria for searching is major, location, name.

The system provides an easy-to-use searching function to help students search information of universities by name, location, major, with exactly information from universities. Students after take the MBTI test, the system will know user's MBTI type and suggests the most suitable major for students, not only major but also universities have highest rating that educating this major.

Student can also read the article of favorite universities, beside that, they can rating, write review and Q&A function can give students more informations about university. With taken informations, student can make a good choice for their future.

In Section 2 we formalize the problem and describe our approach to solving it. Section 3 describes the data structures that need to be set up and the algorithms.

We analyze our algorithm in Section 4. Experimental results and the software used are described in the concluding Section 5.

II. PROBLEM AND SOLUTION PLAN

**We divided into the following main functions:
Search and Suggestions process:**

Problems: Students don't have any system can contain detail and constructed information about 700 universities in Viet Nam. All information are put in article format so that is difficult for students to search the infor that they want to and after searching, student don't know which university is the best or good to keep reading detail information.

Solution: The approach to solving this problem is gives constructed information and search engine help users searching:

- Correlation Recommendation:

This recommendation is the way to suggest university for user. This recommendation based on location of university, train system, score of university. Because there are many universities, it is difficult for users to find similar universities, so our system will suggest universities based on university that user is now viewing.

- **Suitability Recommendation:**

This recommendation is the way to suggest university for user. This recommendation based on result MBTI of user and review about university. Because there are many universities, it is difficult for users can find the university that suitable their personality, so our system will suggest universities based on the MBTI result of user.

MBTI test:

Problem: When the student get ready to study in university, mostly can't decide which career is suitable with their personality, which major they can study well.

Solution: The solution is build a system that allow student to login and take a personality test to know which career is suitable with the MBTI test.

With taken informations, student can make a good choice for their future and the system have the information about MBTI type of student

Rating and Review, Q&A:

Problem: Students are really need informations about university's introduce, major, score and how other peoples, specially is the other students of university think about it, that university is good enough for studying?

Solution:

- The system allow student to login and writing a rating and review about university, and that rating and review of student will help system to recommendation to the students who need the truly information about university.

- The Q&A system is like a mini forum and students can make a thread and other student will answer as they can, and the best answer will recommended if it have most agree/disagree from user

Of course all of the review and question must be accepted by admin.

Favorite and Article:

Problem: If the students want to start studying in an university, they must be searching information and article about that university by themselves.

And they will be lack a lot of article or information about that university if they don't update daily. And maybe the unauthorized and fake article will be readed

Solution: The system allow users to favoriting and following universities, so they can read all article

about that university day by day, the newest, hottest article,...accepted and review by admin. That can provide more information to students who want to studying in that university

III. PLAN IMPLEMENTATION

1. System Architecture

- Our system is developed based on MVC architectural style. We choose this architecture because of the following advantages:
 - We can organize the code better for maintainability, extensibility, and reusability
 - Business logic developers can build the classes, while the UI developers can involve in designing UI screens simultaneously, resulting the interdependency issues and time conservation.

2. Database

- SQLServer
 - SQLServer is very easy to use. With only a few simple SQL statements, you can build and interact with SQLServer.
 - SQLServer is designed to meet even the most demanding applications while ensuring optimum speed, full-text indexes and unique memory caches for enhanced performance.

In order to solve above problems, we apply technologies below, combine them into the algorithms which is described below

■ Correlation Recommendation:

- Get all university has the most similar majors with viewing university
- Get information of university from database
- Find the largest number of similar majors in list S
- Calculate correlation by Pearson correlation coefficient:

$$\text{Correlation} =$$

$$\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} * \frac{n}{t} \right) * 0,6 + a * 0,2 + b * 0,2$$

Call: A is university being viewed by user

B is university to calculate correlation with A

n: number of same majors of B with A

t: The largest number of similar majors in list S

x: score major latest year of A

y: score major latest year of B

a: location point (0-1)

- location of A like location of B: 1 point
- location of A not like location of B: 0 point

b: train system point (0-1)

- train system of A like train system of B: 1 point
- train system of A not like train system of B: 0 point

□ Sort list by correlation descending

■ Suitability Recommendation:

- Get all university has the similar majors with

suggested major based on MBTI type (list S).

- Get review of list S from database
- Find the largest number of similar majors in list S
- Calculate suitability:

$$\text{Suitability} = 0,7 * a + 0,3 * b$$

n: number of similar majors of university with suggested major based on MBTI type
 a: The largest number of similar majors in list S
 b: recommended study university point (0 to 1)
- Sort list by suitability descending

IV. ANALYSIS

In this section, we analysis the performance and functionality of 2 algorithms which mentioned below:

- Correlation Recommendation

The complexity of this algorithm is O(n).

- We execute Correlation Recommendation and receive response after 1.5 seconds.

- Suitability Recommendation

The complexity of this algorithm is O(1).

- We execute Suitability Recommendation and receive response after 0.2 seconds.

V. EXPERIMENTAL RESULTS AND CONCLUSION

Our system has solved the basic problem which is set from the beginning. Users feel convenient by using our system to finding information about universities. The constructed data and flexible search engine give the best experiment to user when try out our system. Users very exciting and satisfying by our recommendation and rating review, there are not any system in VN have this before, so it helped very much by giving appropriate information for their future.

Besides these achievements, our system still exists some limitations:

- The information of very many of university make us hard to collect and update daily, it need more people and budget.
- Information, review, article,... Must be accepted so administrator team must work hard and be available 24/24. It take to much cost to keep the system continue.
- The MBTI test can't prove truely source so can't make users trusted, and it is not only type of personality test, should update more personality test and job test.

ACKNOWLEDGMENT

We would like to thank to Mr. Lam Huu Khanh Phuong for helping us define project scope and improving the functionality.

REFERENCES

- [1] <https://kenhtuyensinh.vn/>
- [2] <https://www.tuoitre.vn/>
- [3] <http://diemthi.tuyensinh247.com/>
- [4] <http://www.maximusveritas.com/wp-content/uploads/2017/07/MBTI-Personality-Type-Test.pdf>

Warning System on Parameters of Devices in the Family:

Deploying warning system for the family in Vietnam

Vo Phuc Hai
FPT University
HCMC, Viet Nam

Ton That Minh Tri
FPT University
HCMC, Viet Nam

Nguyen Phu Ngoc Trai
FPT University
HCMC, Viet Nam

Nguyen Kim Cuong
FPT University
HCMC, Viet Nam

Abstract

Nowadays, home accidents are one of the main cause of injury and even death of people. They are only relating accidents to traffic accident or accidents in outdoor activities. However, as a matter of fact, the place where people regard as safest place home hide many risks: fire house, electric shock, gas and smoke leakage... So a warning system is necessary for home, it help people avoid accident happen in home and protect asset. But in Vietnam, the warning system on parameters is not popular, most of them are deploy in large company, Base Transceiver Station or high important place and very expensive. The big question is whether it is becoming popular in Vietnamese family.

We will provide a system will solve the problems mentioned above. It helps user can check a risk of devices with low installation cost but still work well and suitable for family. We built a system with featured functions: tracking device status, send system warning for user when devices error and remote device control. With solution is developed in Vietnamese family, we provide an easy-to-use and family friendly system. The solution detail will be described in the paper.

Keywords

IoT; CCU; Node processor; RF; encoder; photoregistor; PIR

I. INTRODUCTION

The system will use sensors which mounted on the devices in house and sends parameters of these devices over the internet to report for users via mobile devices. Our system will auto check devices and send an error message to users when problems, errors occur from electrical appliances in house.

We build a system using IOT technology to transform the parameters data to remote sever through internet network. We build app for checking and take warning from system.

II. PROBLEM AND SOLUTION

We have researched many information about current warning system, analyzing their strengths and weaknesses. We find out that most of these system is not a family system, and it is difficult for Vietnamese people to own a similar system. It is usually installed on the BTS or weather station with extremely high value. After that, we find factors and give out solution to adapt similar system, improve and miniature system to suitable for devices in the family. Our three key points of our system are: tracking device status, send warning for user and control device remotely.

At first, we build a system which allows customers to register and login into system. The system will be configured to display list of devices for tracking. The system will be implemented the warning system on parameters of device in the family.

Our solution architecture contains Raspberry Pi for main board, using wireless signal transmission to send data, LCD 5 inch and mobile to access web application communicate with server to control system.

We have 15 problems describe as below:

1. Monitor device with central control unit and node processor

We develop our system with 2 main processor type: central control unit and node processor. Central control unit (CCU) connects with server via internet and node processor via radio frequency (RF). CCU collect data of node processor's sensor and send to server by SignalR Hub. Received data would display on web application on CCU's LCD or user's smart devices such as PC, smart phone.

We develop our system to have main control board Raspberry Pi 3 connect with server and slave board module which contain Arduino, module RF 433.... Main board received data from server and slave board, display data received by web browser using LCD or Mobile phone.

Supervisor: Nguyen Duc Loi, M.E
(Lecturer of IT Department, HCMC FPT University)
Defended topic in front of the scientific council of HCMC
FPT University on December 14th, 2017
Accepted post on December 26th, 2017

2. Central control unit solution: Raspberry Pi 3

We develop central control unit on Raspberry Pi 3 because it has stable enough modules and features for controlling and monitoring home's devices. It has Wi-Fi module to connect to server via internet, GPIO to adding RF module to connect to node processor via radio frequency (RF), HDMI and USB port to connect to touch LCD to show and control device's status in a home, and low power consumption.



Figure 1: Raspberry Pi 3

3. Node processor solution: Arduino Pro mini

In our system architect, node processor collects data of sensor and send to central control unit (CCU) when requested by CCU. We choose Arduino Pro mini for node processor because of these reasons: GPIO enough to connect with device's sensors of all device type, can run stable for a long time, provided by reputable company, low power consumption, low price, and small size.

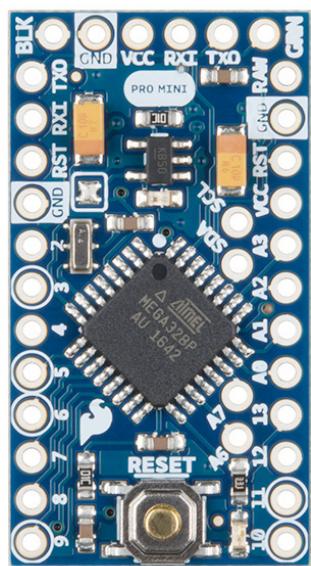


Figure 2: Arduino Pro Mini

4. Wireless communication

RF 433MHz module is used to communicate between Central Control Unit and node processor. We also

use Virtual Wire library to send and receive a string for RF module solution because this library provides specially formatted data, with sync patterns, equal balance of 0 and 1 bits, and error checking for RF communication. Using RF 433MHz module also reduces changes to be interfere with others Passband frequency.

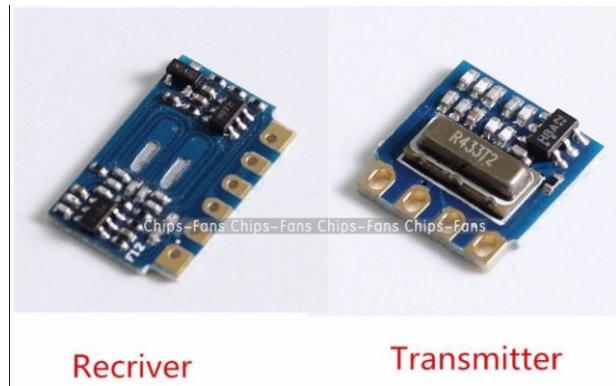


Figure 3: Transmitter and Receiver 433Mhz module

5. Send warning to user: using SMS API

We using web browser to display information of device and show warning. In addition, we use SMS to send warning messages for user. If any parameters exceeded the allowed level, the system would send warning message and suggested solution to user's phone. We use a SMS API of eSMS, a Vietnamese company, to send warning message to user. This API allow sending SMS with content length up to 160 characters, sending to multiple receiver as a same time, and sending SMS with registered brand name – SMS with sender name is our own brand name – to anti phishing.

6. Measure electricity solution: Pzem-004T module

To apply a warning system to traditional house, first of all, we must monitor electricity source of the house. Therefore, we need a module to measure voltage, current, power of a house to make sure that all devices of a house are operating with proper voltage, current and user can monitor total power of all devices of their house. But we have a problem that all of us don't have an embedded knowledge enough to make a module for 220V electric as we need. Luckily, we found pzem-004T module which meet our requirements. This module transfer measured data to node processor by UART protocol.

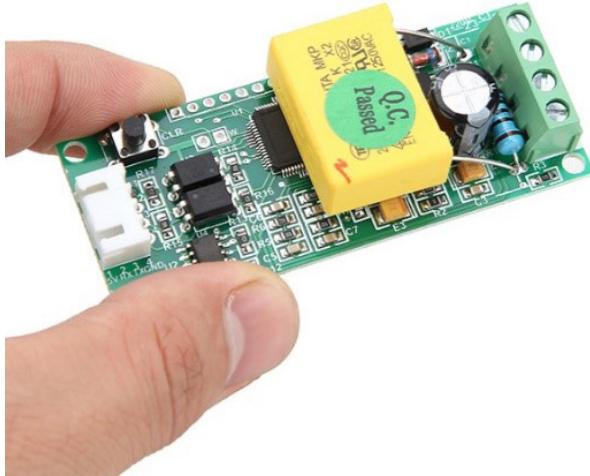


Figure 4: Pzem-004T module

7. Check the electrical of device

With popular electric devices which have 2 states: on and off, our warning system have to make sure that:

- In on state of toggle switch, the device must be supplied electricity source.
- In off state of toggle switch, the device must be disconnected to power.

Out of two above case, our system would send warning to users about this case to protest their device. To solve this problem, we have a simple solution: using a AC to 5V DC adapter. When the device is supplied electricity source, DC output of adapter will send HIGH signal to node processor. In otherwise, LOW signal is sent to node processor. Depend on level of adapter's signal, our system can monitor reality state of devices.

8. On/off switch device by wireless

Many switches used in our house are a traditional single-pole switch, which are just a simple switch connected from source to load, also the clicking sound is somehow annoy people. We replace these switch to our own RF switch, which is a combination of RF module and relay control by the CCU. It is also feature touch sensor for convenient control device.



Figure 5: Relay module

9. Measure Spin solution: encoder module

Nowadays, there are many devices with electric motor in a traditional house such as fan, pump, ... So, measuring speed of motor is necessary to make sure that the electric motor is in good condition, and avoid over heat resulting in damage the device. Encoder module is our solution for this situation. It is mounted in electric motor and connect to node processor of device to monitor motor's speed based on device's state.

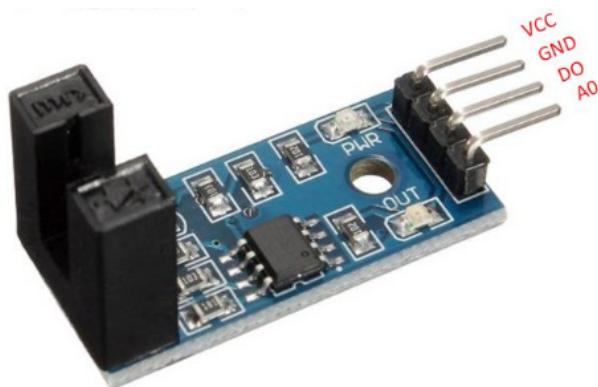


Figure 6: Encoder module

10. Measure Temperature of device solution: LM35 sensor

Almost the fire at home is result of electric devices overheat. Because of that, our system must also monitor device's temperature to avoid overheat situation by using LM35 sensor which is IC from Texas Instrument - famous chip manufacturer in American. It is mounted in device and connect to node processor of device to measure device's temperature of device.

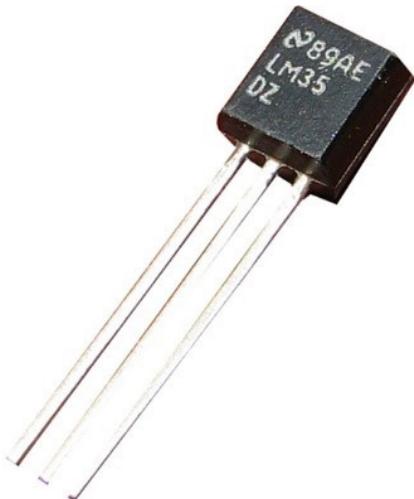


Figure 7: IC LM35

11. Detect light solution: photoresister

Electric lights are basic devices of each family. Sometime, the electric light is not bright when user turn on the light. So, a photoresister mounted on the light is using to monitor the electric light of our warning system. A warning message will be send to user's phone when there is any problem with the electric light.

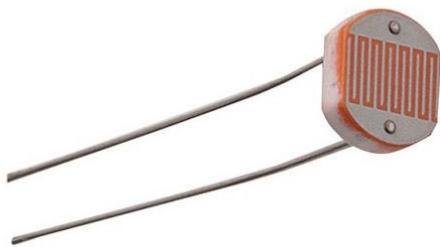
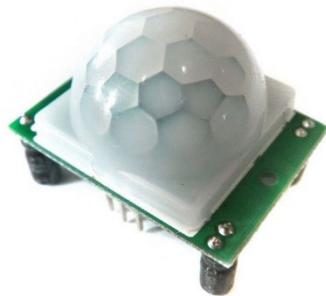


Figure 8: Photoresistor

12. Security an area solution: Passive Infrared Sensor PIR HC-SR501

There are many solutions for security problems in the house, mostly to prevent theft or unwanted trespassing. Camera is the most practical way to monitoring the area, but still has blind spot. Passcode, PIN code, finger print and magnet door is just an alternative solution. For that reason, PIR or passive Infrared Sensor come in handy. This sensor detect motion from human being, whenever a person get in the sensor detection area, a signal feedback to the owner's property CCU.



**Figure 9: Passive Infrared Sensor
PIRHC-SR501**

13. Measure environment temperature solution: DHT22

Some devices like refrigerator, television, personal computer requires specific room's temperature and humidity for their operation. Depending on device in the room, we will configure our system to monitor environment temperature and humidity of the room with DHT22 module – temperature and humidity sensor – to make sure that user's devices are operated in good condition. DHT22 module is also mounted in ceiling of the room and connect to room's environmental node processor.

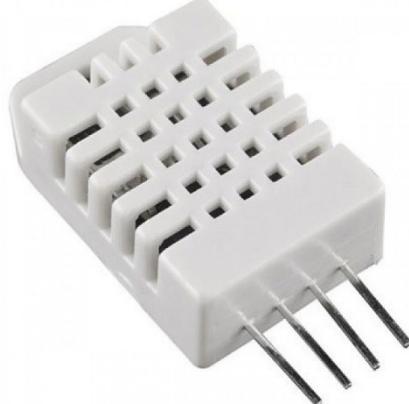


Figure 10: DHT22

14. Fire detection & smoke detection system

Our warning system not only protect an electric device, but also protect a house of user. Fire and Smoke detector is made by ourself on our knowledge about electronic, and how a general fire and smoke detector works. Fire and Smoke detector is mounted in ceiling of the room and connect to room's environmental node processor to detect fire and smoke in the room.

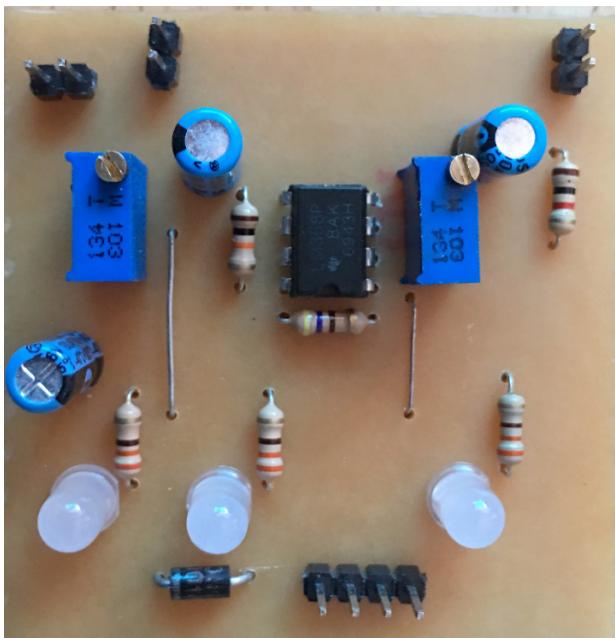


Figure 11: Self-design PCB for fire detection & smoke detection system

15. Gas detection solution: MQ-2

To protect your house out of risk of the fire, we use MQ-2 module – gas detection sensor – as our solution for our warning system. MQ-2 module is also mounted in ceiling of the room and connect to room's environmental node processor to detect gas in the room.

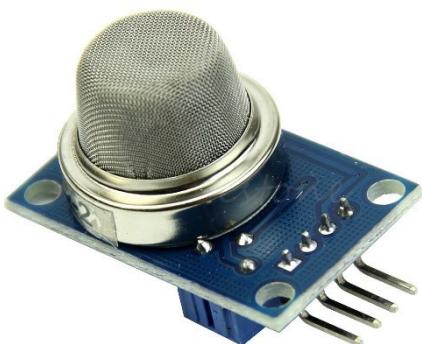


Figure 12: MQ-2 module

III. PLAN IMPLEMENTATION

1. Monitor device with central control unit and node processor

We build a web application for user to control and monitor their house at anywhere, on any platform. At the first time, the user will login to system with the given account. After login successfully, the application will connect to socket server to get the house's status of devices, which is configured before, in real time. User uses the web application on your smart device such as central control unit, laptop,

and smart phones to control and monitor devices as well as get notification about new warning of their devices. The devices in a house are controlled and monitored by 2 main components: Central control unit and node processor.

Central control unit (CCU) run a python application to get sensor's data and control device's state. Firstly, CCU boot up and run the configured python application. The application will connect to API server to get information of all devices and sensors in a house, token to sign in to web application with account's information which configured before. Next, CCU will connect to socket server SignalR to configure real time connection. Then, CCU open web application for user to control and monitor the house. CCU connect to node processor which mounted in devices through RF 433MHz module via Radio Frequency (RF) and require node processor send sensor's data back by sending a package contain deviceId of mounted device. A deviceId is a string which has 5 characters such as: "htc8", "mtia8", ... Sensor's data pushed to server then pushed to user's web application after processed by server. When user remote device on web application, a remote command sent to server. Then CCU will receive command from server, then it will send command to node processor which mounted in the remoted device.

Node processor is an embedded board which programed by C language. When running, node processor collects data of sensors in real time, saves to variables, and waits for command from CCU. When node processor receives a package which contain itself deviceId, which configured before, through RF 433Mhz module via Radio Frequency (RF), it will send sensor's data from variables and current device's state back to CCU immediately. When node processor receives a package which contain itself deviceId and 1 more bit to control device's state such as: deviceId is "htc8", control command is: "htc80" to turn off device, "htc81" to turn on device.

2. Wireless communicate between CCU and node processor

RF 433MHz module is mounted at both Central Control Unit and node processor. The RF transmitter module is supplied 9VDC power source for better performance. Setting up RF module for communication with following steps: Firstly, Virtual Wire library is imported to application. Secondly, setting transmit pin and receive pin for RF transmitter module and receiver module. Thirdly, setting baud rate for communication. Notice that RF 433 modules can communicate together only when

they are set up same baud rate. Finally, starting received service as a listener. To send message via RF, `vw_send()` function is called. It takes about 50 to 200 milliseconds to make sure that the message is successfully sent.

3. Send warning to user

We use SMS API of eSMS – a Vietnamese company – with following setting steps: First, login to eSMS website with registered account to get APIKey and SecretKey. Second, configuring request URL for sending GET request to eSMS API server. The request URL like this:

http://rest.esms.vn/MainService.svc/json/Send-MultipleMessage_V4_get?Phone=01644211555&Content=TesMessage?ApiKey=XXXXXX XXXXXXXXXXXXXXXXXXXXXXXX&SecretKey=XXXXXXXXXXXXXXXXXXXXXXX&SmsType=4 with “Phone” is phone number of receiver, “Content” is content of message, “ApiKey” and “SecretKey” are the keys which provided by eSMS, “SmsType” is type of message. Each time API is called, it returns a result in Json type that contains state of sent message is successful or not.

4. Measure electricity solution:

To measure electricity of family with Pzem004T. We configure the input like figure below:

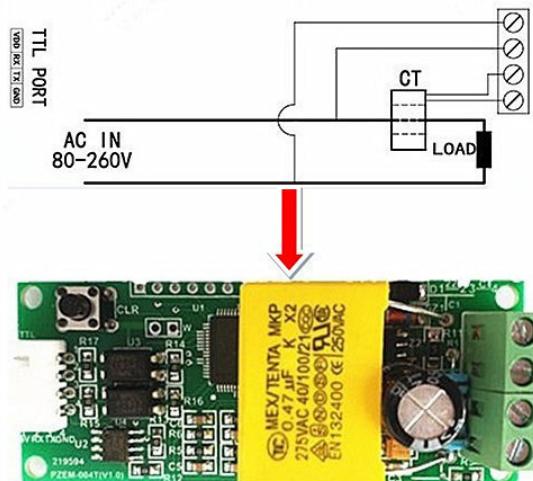


Figure 13: Configure input of Pzem004T

Pzem004T communicate with node processor via UART protocol. So, to connect this module’s pin to node processor’s pin, we follow below table:

Table 1: Pzem004T pin and node processor pin mapper

Pzem004T pin	Node processor pin
VDD	VCC
RX	TX
TX	RX

GND	GND
-----	-----

Before measuring, we need to set up this module with node processor. First, PZEM004T library is required. Next, we configure Serial port of node processor for UART protocol. `Voltage()` function of library is called, it returns voltage of electricity. `Current()` function returns current of all load devices. `Power()` function returns total power of all load devices. It takes 2 seconds for each measurement.

5. Check the electrical of device

We use AC to 5V DC adapter to check the electrical of device. This module like phone charger, its output is 5V DC matching with High level of digital input of node processor when its input is in range 100-240V AC – it means that mounted device has power source. When mounted device doesn’t have power source – it means input of this one is 0V AC, its output is 0V DC matching with Low level of digital input of node processor. To connect this module to node processor, we follow below table:

Table 2: AC to 5V DC adapter pin and node processor pin mapper

AC to 5V DC adapter	Node processor pin
VCC	General digital input
GND	GND

After connect this module to node processor, we configure pin mode of connected pin is INPUT. We use function `digitalRead(configured pin number)` to read output state of adapter module.

6. Measure motor’s speed

Encoder module is used to measure electric motor’s speed. Before measuring motor’s speed, we need to set up this module. First, we mount a round plastic with holes on axle of the motor. Then, we put this module below a plastic like below figure.

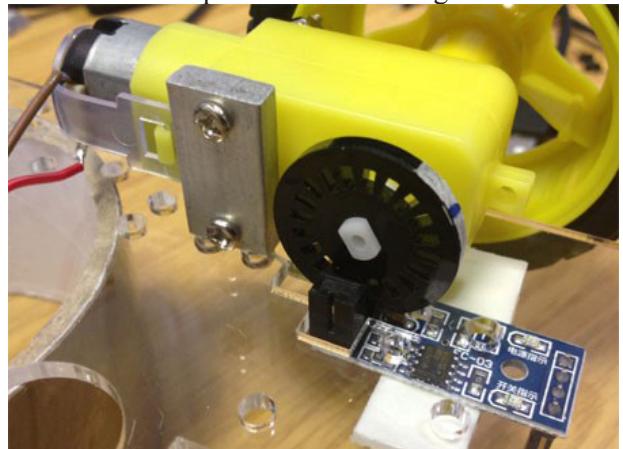


Figure 14: Setting up encoder module

Next, we will connect encoder module to node

processor. This module has four pin: VCC, GND, DO- digital out, AO – analog out. However, we just need 3 pins: VCC, GND and DO. To connect this module to node processor, we follow below table:

Table 3: IC LM35 pin and node processor pin mapper

IC LM35	Node processor pin
VCC	VCC
DO	Interrupt digital pin
GND	GND

After connect this module to node processor, we configure pin mode of connected pin is INPUT and attach interrupt function to interrupt digital pin in FALLING level signal condition to calculate motor's speed.

7. Measure Temperature of device

IC LM35 is used to measure device's temperature by mounted it on device's surface. This IC has 3 pin: Vs - supply voltage, GND – ground, Vout – output voltage. To connect this module to node processor, we follow below table:

Table 4: IC LM35 pin and node processor pin mapper

IC LM35	Node processor pin
Vs	VCC
Vout	A01
GND	GND

After connect this module to node processor, we configure pin mode of connected pin is INPUT. There is a function to calculate mounted device's temperature from analog signal of IC LM35.

8. Detect light

Photoresistor module is used to detect if mounted lamp is bright. Before using this module, we need to mount a photoresistor of module on the illumination surface of the lamp. This module has 3 pin: VCC, GND, DO – digital out. To connect this module to node processor, we follow below table:

Table 5: Photoresistor module pin and node processor pin mapper

Photoresistor module	Node processor pin
VCC	VCC
DO	General digital pin
GND	GND

After connect this module to node processor, we configure pin mode of connected pin is INPUT. We use function digitalRead(configured pin number) to read output state of module.

9. Security an area

Passive Infrared Sensor PIR HC-SR501 is used to detect motion in a room. Before using this module, it need to mounted on the ceiling like below figure for best performance.



Figure 15. Mounting a PIR on the ceiling

This module has 3 pins: VCC, GND, DO – digital out. To connect this module to node processor, we follow below table:

Table 6: HC-SR501 module pin and node processor pin mapper

Passive Infrared Sensor PIR HC-SR501	Node processor pin
VCC	VCC
DO	General digital pin
GND	GND

After connect this module to node processor, we configure pin mode of connected pin is INPUT. We use function digitalRead(configured pin number) to read output state of module.

10. Measure environment temperature and humidity
DHT22 is used to measure environment temperature and humidity in a room. Before using this module, it need to mounted on the wall for the most accurate result. This module has 4 pins: VCC, GND, DATA – digital out, NULL. However, we just need 3 pins: VCC, GND and DATA. To connect this module to node processor, we follow below table:

Table 7: DHT22 module pin and node processor pin mapper

DHT22	Node processor pin
VCC	VCC
DATA	General digital pin
GND	GND

Before measuring, we need to set up this module with node processor. First, DHT library is required. Next, we configure connected pin for DHT library. readHumidity() of library function returns humidity

inside the room. `readTemperature()` function returns temperature inside the room.

11. Fire detection & smoke detection system

Fire & smoke detector module is used to detect fire and smoke in a room. Before using this module, it need to mounted on the ceiling like below figure for best performance.

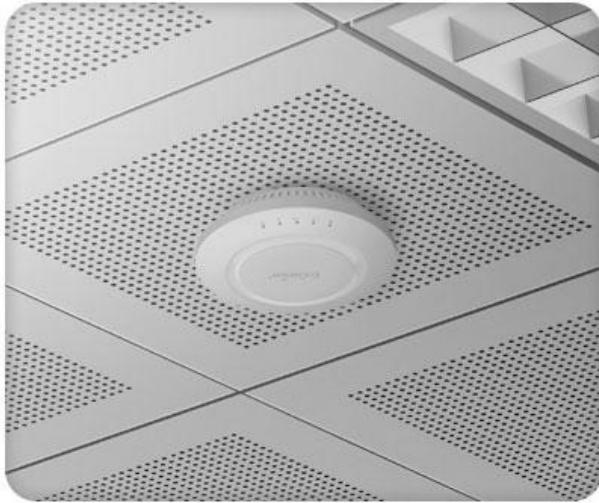


Figure 16: Fire and smoke detector mounted on the ceiling

This module has 4 pin: VCC, GND, DO1 – digital out of smoke, DO2 – digital out of fire. To connect this module to node processor, we follow below table:

Table 8: Photoresistor module pin and node processor pin mapper

Fire & smoke detector module	Node processor pin
VCC	VCC
DO1	General digital pin 1
DO2	General digital pin 2
GND	GND

After connect this module to node processor, we configure pin mode of connected pins is INPUT. We use function `digitalRead(general digital pin 1 number)` to read output state of smoke detector, `digitalRead(general digital pin 2 number)` to read output state of fire detector.

12. Gas detection

MQ-2 module is used to detect gas in a room. Before using this module, it should be mounted near gas source for best performance. This module has four pin: VCC, GND, DO- digital out, AO – analog out. However, we just need 3 pin : VCC, GND and AO. To connect this module to node processor, we follow below table:

Table 9: MQ-2 module pin and node processor pin mapper

MQ-2 module	Node processor pin
VCC	VCC
AO	General analog input
GND	GND

After connect this module to node processor, we configure pin mode of connected pins is INPUT. There is a function to check if gas is leakage from analog signal of MQ-2 module.

IV. ANALYSIS

In this scope, we implement Central Control Unit of our system on main board Raspberry Pi 3 is a credit-card-sized computer which has 1.2GHz 64-bit quad-core ARM Cortex-A53 CPU, 1GB RAM, integrated 802.11n wireless LAN, 4 USB ports, Ethernet port and 40 GPIO pins (<https://www.raspberrypi.org/help/faqs/>). After doing system test we have some analyze and risk assessment below:

1. Central Control Unit Raspberry Pi 3

Our Central Control Unit uses Raspberry Pi 3 made in England. This board runs Raspbian OS which based on Debian optimized for the Raspberry Pi hardware. Almost 100% of the time, this module work fine. The Wi-Fi connection is stable, but Wi-Fi range is not really wide. Best range for Wi-Fi connection is about $7m \pm 2m$. There are some ways to increase Wi-Fi range of Raspberry Pi but it requires many technical skills about anti-jamming and power resources. Because of our limited resources, we put Wi-Fi access point near this module for best connection.

Raspberry Pi can connect to internet via wire, but it is not convenient when Wi-Fi connection available and the result doesn't make sense.

2. Transmission data using RF communicate

We use third party RF modules: H5V3M for receiving and H34A for transmitting which made in china. Both of them have working frequency at 433MHz. This module is unstable because of losing package. Probability of lost package depends on supply voltage for transmitter, distance between transmitter and receivers, and size of package. Each package contains about 4 to 50 characters, supply voltage for transmitter from 5VDC to 12VDC, supply voltage for receiver is 5VDC.

Table 10: RF communication in about 7m distance test in 100 times

Supply Voltage	Character	Rate of lost package
5VDC	5	5%

5VDC	20	35%
5VDC	50	78%
9VDC	5	2%
9VDC	20	12%
9VDC	50	34%
12VDC	5	0%
12VDC	20	5%
12VDC	50	13%

3. Real time communication

SignalR technology is a modern Web Socket of Microsoft. We have tested our solution with home's Wi-Fi connection (Ping 7ms, Download: 21.5Mbps, Upload: 15Mbps) for many time and the result is very good. Connection is stable. Delay time is usually under 300 milliseconds, no lost package found and disconnection does not occur in almost case.

4. Send warning to user

SMS API of eSMS is a service of Vietnamese Company, so we guest that the delay time would be low. In our reality test, we have send SMS to our phone by sending GET request to their server in 20 times with account balance is available. Although the delay time is not as our expected - about 2 to 5 seconds, the lost message does not occur. So, it meets requirement of research scope.

5. Measure electricity solution

Our solution uses PZEM004T module of PeaceFair company – a Chinese company. This module is stable and high precision. We have tested this module with home's power source in 10 time and compared with multimeter. In almost case, the output of this module is similar the result of multimeter, see below table. A disadvantage of this module is a response time. It takes 2 second for each measurement.

Table 11: Comparison between output of PZEM004T module and result of multimeter

	Output of this module	Result of multimeter
Home's power source 1st time	232.0V	232.6V
Home's power source 2nd time	230.6V	231.3V
Home's power source 3rd time	226.2V	226.7V
Iron's current	8.10A	8.212A
Fan's power	36W	N/A
15W Bulb's current	0.06A	0.067A

6. Check the electrical device

AC-5V DC adapter is simple, low price and stable. It's work correctly in almost our test cases. Although it cannot measure exactly voltage electrical of device, it meets requirement about check the electrical of device. A disadvantage of it response time. It take 1 -2s to detect current electrical state of device.

7. Measure motor's speed

The encoder module which we use is no brand module with low price (under \$1). We guest that this module made in China. When we compare output of this module with RPM meter in reality test, this module is low accuracy. Error of it is approximate 15%. This module is unstable. We will improve this hardware for this function in next version.

8. Measure Temperature of device

The LM35 temperature sensor are precision integrated – circuit temperature sensor, with an output voltage linearly proportional to Centigrade temperature. We guest that this module made in China. With error of it is approximate 2%, this module is quite stable and good performance for measure devices need correct parameters.

9. Detect light

Photoresistor module which we use is no brand module with low price. We guest that this module made in china. However, the quality of it makes us surprise. It responds with no delay, and work correctly in almost case of our test.

10. Security an area

Passive Infrared Sensor PIR HC-SR501 – a popular and famous sensor – is widely used for detect motion in empty area. So, we think that this module should work perfectly. In our reality test, this sensor detects both slow and fast motion of our test cases. Almost 100% of the time this sensor works fine.

11. Measure environment temperature and humidity
DHT22 Temperature Sensor – a popular temperature with a good price (from 4\$ - 5\$) and easy to retrieve data by 1 wire communication (1 – wire digital interface). This module has accuracy humidity $\pm 2\%$ RH and temperature $<+0.5$ Celsius. So, we think that this module work perfectly for measure environment. The data responds with no delay in almost case of our test.

12. Fire detection & smoke detection system

We use our self-design PCB for fire detection & smoke detection system.

This module can detect fire and smoke in 70% our test cases. It has high sensitivity, can adjust level of sensitivity by rheostat but it's not really stable. We have found the reason for 30% failure in our test cases, it's caused by jamming. There are some ways to make it stable for a long time but it requires many

technical skills about anti-jamming. Because of our limited resources, we will improve this module for better performance in next version.

4.13. Gas detection

MQ-2 gas sensor which we use is a popular module with low price. It is useful for gas leakage detecting (in home and industry). It can detect H₂, LPG, CH₄, CO, Alcohol, Smoke and Propane. Based on its fast respond time (using digital interface). Measurements can be taken as soon as possible. Also the sensitivity can be adjusted by the potentiometer. With these advantages, almost 90% of time this sensor works fine.

V. EXPERIMENTAL RESULTS AND CONCLUSION

We have done system test in a real house with three rooms of house: 1 kitchen, 1 bedroom, and working room.

The result of monitoring the house we received is described as below:

- Intrusion warning detector Room: 99% detected.
- Environment monitoring Room: 87% passed.
- Devices parameter monitor and control Room: 90% passed.

In conclusion, we create a system for safety of people, and assets. In reality, we found that there is no system like our system for traditional house on the market. Our system is very convenient to install and use. The price is also affordable allowing people to live more comfortably with new life style. A warning system for a house will be new trend beside smart home, support for smart home to bring safety, assurance, happiness to every single family.

ACKNOWLEDGMENT

First and foremost, we would like to express our greatest gratitude to Mr. Loi, our supervisor, for his support and guidance during this project, especially in fulfilling project requirements and improving project's quality.

In addition, we would like to send our special thanks to all FPT University Lecturers who have taught and guided us from the very beginning. This has been a great instruction in this strenuous journey.

REFERENCES

- [1] Datasheet sensors and communication with arduino
<http://arduino.vn>
- [2] Raspberry Pi 3 model B datasheet available
<https://www.raspberrypi.org/products/raspberry-pi-3-model-b> .
- [3] Send and receiver data by RF 433Mhz
<http://randomnerdtutorials.com/rf-433mhz-transmitter-receiver-module-with-arduino/>
- [4] Arduino Pro mini
<https://www.engineersgarage.com/electronic-components/arduino-pro-mini-pinout>
- [5] Control devices
<https://openhomeautomation.net/control-a-relay-from-anywhere-using-the-raspberry-pi>

Papers of Economic Sector

Analyzing the Consumer Satisfaction for an Internet Service Provider

Vo Thi Tuyet Nhi

FPT University

HCMC, Viet Nam

Le Thi Diem Xuan

FPT University

HCMC, Viet Nam

Vo Quang Sang

FPT University

HCMC, Viet Nam

Pham Anh Duc

FPT University

HCMC, Viet Nam

Abstract

The Internet has become an important aspect of everyday life in modern societies. The development of the Internet is one of the important factors that makes it easy and convenient to convey information to other parts of the world. According to Rajan Anandan, vice president and CEO of Google Southeast Asia and India, Vietnam has the most dynamic Internet market in the world because it is the only market that has more Internet users than people not used (2016). Obviously, surviving and expanding market share in this competitive market are the major challenges for the internet service provider. The key to maintaining the market share is customer satisfaction, so the purpose of this study is to analyze the customer satisfaction of the Internet service provider. Factors in this research include Tangible, Reliability, Responsiveness, Assurance and Empathy. We conducted the questionnaire with 300 - 400 respondents living in Ho Chi Minh city and using the Internet. Data has been analyzed by using SPSS software (version: 22). Through survey and analysis, it proved that, there are 4 main factors including Tangible, Reliability, Responsiveness and Assurance were accepted. This report can analyze the influence of these factors on the overall satisfaction of customer to Viettel internet service provider. From the outcomes of the report, the authors propose some managerial implications to corporations to increase customer satisfaction by improving the quality of service better.

Because of the necessity of maintaining customer satisfaction, the researchers do this study to analyze the factor that affects to customer satisfaction in Vietnamese Internet Service Provider.

This research examines the application of theories in by developing a proposed research model. And then the authors use some research methods analyzing data to find out the practical meaning of Vietnam internet market.

We adopted a structured questionnaire-based survey comprising 32 questions and distributed to 350 responders within Ho Chi Minh City. On the other hand, the result of regression analysis proved 5 factors: Tangibility, Reliability, Responsibility, Assurance, Empathy have significant positive influence on customer satisfaction for Viettel ISP. In the other words, it is worthwhile for companies to use their resources by focusing on customer satisfaction.

Keywords

Customer Satisfaction

I. INTRODUCTION

Survey Respondents are people from individuals or organizations who are using internet services from Viettel provider in Ho Chi Minh City. This research focuses on consumer's behavior and their satisfaction in using internet service of provider Viettel in Ho Chi Minh City. The survey respondents who have used or are using internet services such as individuals or organizations. The number of respondents is about over 300 samples in four weeks in October 2017.

This research focuses on customer satisfaction of Viettel internet service provider. Viettel provides the services how to satisfy customers; as well as customers want to improve and upgrade the service package of the network from Viettel. The research will be conducted in 24 districts of Ho Chi Minh City where the number of internet users is larger and more popular than in other places.

Chapter one guides the reader with all the information needed for the research include the proposal by characterizing subject definition, explore foundation, common sense issue, inquire about goal, and approach.

This chapter clarifies the hypothesis, review theories, conclusion from the previous study, selected theoretical model and build framework. In addition, it also contributes for this study to be more

Supervisor: Vo Minh Hieu, MBA. (Head of Business Department, FPT University, HCMC FPT University)

Defended topic in front of the scientific council of HCMC FPT University on December 16th, 2017

Accepted post on December 26th, 2017

reliable by mentions a specified examination about interactions between customer loyalty and relevant factors.

This chapter determines the research method, data collection methods, questionnaire's development, data analysis methods and limitations of the research project. This section shows Research results and data analysis. Data analysis methods are used including descriptive analysis, Exploratory Factor Analysis (EFA), testing the Reliability of Scale by using Cronbach's Alpha, Confirmatory Factor Analysis (CFA). Those operations are utilized by SPSS 22. This part displays the study conclusion with discussing the main findings and gives recommendations. The limitations of the paper and suggestions for future research are also mentioned.

II. THEORETICAL BASIC AND LITERATURE REVIEW

1. Theoretical basis and analysis framework

Parasuraman et al., (1985), developed a model of service quality after carrying out a study on four service settings: retail banking, credit card services, repair and maintenance of electrical appliances, and long-distance telephone services. The SERVQUAL model represents service quality as the discrepancy between a customer's expectations of service offering and the customer's perceptions of the service received (Parasuraman et al., 1985). They identified 97 attributes which were condensed into ten dimensions. They have an impact on service quality and were regarded as the criteria that were important to access customer's expectations and perceptions on delivered service (Kumar et al., 2009, p. 214).

The SERVQUAL scale has been proven to be one of the best ways to measure the quality of services provided to customers. This service evaluation method has been proven consistent and reliable by some authors (Brown et al., 1993). They held that "when perceived or experienced service is less than the expected service. It implies less than satisfactory service quality". It means that when perceived service is more than expected service, which service quality is more than satisfactory (Jain et al., 2004, p. 27). This makes it best fits the evaluation of service quality from the customer perspective.

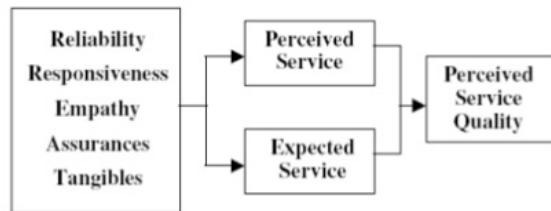
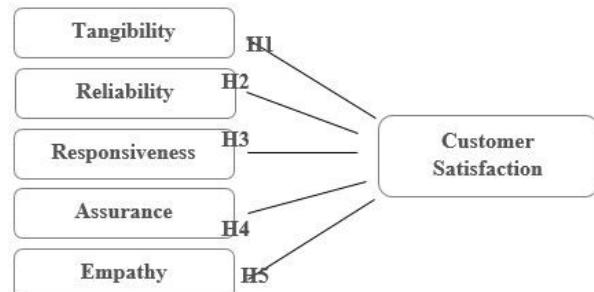


Figure 2.1: Measuring service quality using SERVQUAL model (Kumar et al, 2009)



**Figure 2.3: Hypothesis model
(source: by author)**

2. Literature review

Parasuraman, Zeithaml & Berry, 1985. The five main dimensions in SERVQUAL are tangibility, reliability, responsiveness, empathy, and assurance:

Tangibility

This component in the SERVQUAL Model is about the appearance of physical evidence. As we can see that how the stores can attract the customer such physical facilities, equipment, and appearance of the officer in front of the stores. Continuing into the office, how they custom their office to make their customer feel comfortable and the office needs to be the most convenient for the customer to find the services that they need. Viettel invests in their equipment, especially computer systems to help the staff can do their job fast and create comfort for the customer, that is a good example. According to Parasuraman et al., in the year of 1988, they showed 4 questions to determine tangible dimension:

TA1 Grocery stores have up-to-date equipments.

TA2 Physical facilities are virtually appealing.

TA3 Employees are well dressed and appear neat.

TA4 Physical environment of the grocery store is clean.

Reliability

It was defined as ability to perform the promised service dependably and accurately. Clearly, the firm is able to give and do their promise which convince their customer in the short time quickly and exactly ways.

At the first time let they reach their need and serve their people are included at the power of company, this is the very important thing which part of

customer's servicing. The reliability also enhances to contribute services when promised and conserve the error records without payment. With any companies they can't avoid error which happens after customers purchase and experience, the company always give them warranty which make they more believe that they are protected by policy as promise from company. These following examples show the factors help reliability dimension increase: The staff solves customers' problems encountered that have been guarantee to them on time; the sellers serve their guests as firm promised and punctual. In this term, there are 5 problems that we have to discuss to the guest

RL1 When they promise to do something by a certain time, they do it.

RL2 When customer has a problem, they should show sincere interest in solving the problem.

RL3 Grocery stores perform the service right the first time.

RL4 They provide their services at the time they promise to do so.

RL5 Grocery stores keep their records accurately.

Responsiveness

This factor point out of the willingness or readiness of employees to provide service, timeliness of service such as mailing a transaction slip immediately, calling the customer back quickly, giving prompt service.

The models mentioned above focused on the qualitative research more than quantitative research which is empirically and psychometrically tested. Parasuraman et al. (1988) developed SERVQUAL which is an advanced model for measuring service quality. In SERVQUAL model, there are 5 dimensions and 22 items presented in seven-point Likert Scale. They measured especially functional service quality through empirical studies in banking, credit card, repair and maintenance, and long-distance telephone services.

RN1 Employees make information easily obtainable by customers.

RN2 Employees give prompt services to customers.

RN3 Employees are always willing to help customers.

RN4 Employees are never too busy to respond to customers requests.

Assurance

The courtesy, knowledge, ability of employees to inspire trust and confidence. In the old SERVQUAL model of Parasuraman et al.(1985), competence, courtesy, credibility, and security are four important elements to make the trust of the customer.

Competence indicates a cluster of related abilities

of knowledge and skills that employee to interact with customers in a wide variety of situations. It requires the staff to be able to convey the message to the customer fully and easily understood. Courtesy means polite behavior that shows respect for other people; that will bring the satisfaction of guests when the employee show politeness, professionalization, respect, regard, and friendliness with the customer. Credibility means the quality that service or behavior from the employee has that makes the customer believe or trust them. Security is an integral part of creating trust from customers, as well as promoting customer loyalty to the company's products.

AS1 The behaviour of employees instil confidence in customers

AS2 Customers feel safe in their transactions with the employees

AS3 Employees are polite to customers.

AS4 Employees of grocery stores have knowledge to answer customers' questions.

Empathy

The last component of SERVQUAL model is empathy. Empathy is about showing interest and awareness of what the customer needs, the feeling of experience and the use of customer service. In that component is a combination of Access, Communication, and Understanding the customers. Access is a way of the provider connecting with the customer and vice versa.

The ability to connect as convenient and fast will positively impact positively on customer satisfaction. Communication not only makes conversation to customers by using language that the staff needs to listen carefully to what customers need, what they want to use. They need to control their feeling, their language to talk with many different customers, then they have to explain to the customer if the customer does not understand anything. Understanding customers is the goal and the purpose of all organizations and businesses. It also demonstrates that the supplier is also listening to comments from customers.

EM1 Grocery stores give customers individual attention.

EM2 Operating hours of grocery stores are convenient to customers.

EM3 Employees of grocery stores give customers personal service.

EM4 Grocery stores have their customers' interest at heart.

EM5 Employees of grocery stores understand the specific needs of their customers.

III. METHODS & DATA

1. Data

Sampling methods: non-probability sampling method

Research object: People from 18 to over 25 years old who are using website that market digital products and services in HCMC

Sample size: 396 samples included:

350 samples are in good range for researching Viettel Internet Service provider

30 samples for pilot test (Tabchinick & Fidell (1996) and Comrey and Lee (1992).

Data collection: Questionnaire survey

2. Methods/Framework

- Descriptive statistics: Descriptive statistics provides basic summaries of sample and scale in order to create a foundation for every data quantitative analysis
- Cronbach's Alpha: Cronbach's Alpha ratio is reliability ratio used to estimate correlation scale between observed variable-pairs
- Exploratory Factor Analysis (EFA): Exploratory Factor Analysis (EFA) is a statistical technique which is used to reduce observed variables to a smaller set but still maintains almost information content and statistic meaning of the initial set of variables (Hair et al., 1998).
- Regression Analysis: Regression analysis is one statistical method which is used to evaluate the strength of relations between one or more dependent factors and one or more independent factors.

IV. RESULT & FINDING

- Customer satisfaction in using the Internet that Internet market was affected by the for 5 key factor include Tangibility, Reliability, Responsiveness, Assurance and Empathy.
- The finding does meet the initial expectation that accepts 4 factors.
- Different when compare between satisfaction and service quality suggests different methods for each model to modified and raise the customer satisfaction.

V. CONCLUSION AND POLICY

IMPLICATION

Based on the results of the research team from Chapter 4, the data sheet shows that there are four factors that affect the customer satisfaction of the Internet service provider in the case Viettel provider. Four factors are Tangible, Reliability, Responsiveness and Assurance. In addition, the indicators show that these 4 factors have a great impact on customer satisfaction. Based on these data of factors and the situation of Viettel in practice, the authors of the comments and recommendations below:

VI. TANGIBILITY

As we know that the beta is 0.279 which proves that factor has the effect to the customer satisfaction of Viettel internet service provider but not as strong as others. Because the use of the internet and the registration can be registered at home or company, this factor does not make much impact on customers. According to Sagher and Nathan (2013), they found that the relationship between tangibles and customer satisfaction was not significant enough. However, Viettel still should to focus on this element that makes user prefer to increase customer satisfaction in the near future.

Recommendation

Tangibility represents physical facilities that facilitated the process of service provision which include staff members, equipment and printed materials. Visual images plays a very important role in creating customer satisfaction because it helps create a visual impression that attracts customers, creates an impression of quality of service. Customers always want to work in place with modern equipment and decoration appropriate, impressive (Culiberg & Rojsek, 2010). In additions, based on a study by Nabi (2012), which he concluded that customer want a store has a beautiful location, convenient and modern equipment. That helps create customer satisfaction.

VII. RELIABILITY

According to the multiple regression analysis, beta of perceived usefulness is 0.201 that illustrates this factor has the effect to customers satisfaction of Viettel internet service provider but not strong. Said another way, this element contributes a small part to the Satisfaction of customer. However, This dimension is very important of service in conventional sectors (Parasuraman et al., 1988). Therefore, Viettel should focus on this factor to improve customer satisfaction.

Recommendation

Reliability includes security, privacy and assurance. Security is actually a freedom from the doubts, risks and dangers. It is very important for a company. In internet services, customers are more careful about transactions (Ranganathan, 2002). Privacy means that safe methods are adopted. Every customer is concerned about the security of information and data security. In addition, customer privacy is extremely important to increase customer satisfaction. Therefore, the company should ensure the safety of the services that it provides (Black, 2001).

VIII. RESPONSIVENESS

Based on the data in Chapter 4, the number Beta

of Responsiveness is the highest of the four factors ($\beta = 0.355$). This reflects the Responsiveness is the factor greatest influence on customer satisfaction and shows Customer satisfaction about the enthusiasm factor of Viettel employees. Thus, Viettel needs to focus on this element that make user prefer to increase customer satisfaction in the near future. In detail, the factors include Viettel's employees give prompt services, make information easily obtainable and always willing to help customers creating customer satisfaction for Viettel. Therefore, Vietnam should continue to maintain and develop this factor to increase trust from customers.

Recommendation

According to Balunywa (1995), Responsiveness is an important factor that enhances speed of processing information influence positively on customer satisfaction. Therefore, the company needs to pay attention to the full, always ready and be easy offered when the customer needs. Customers do not like to wait long, so they always want their request to be met the fastest (Ruby et al., 2012). A customer can be becoming loyal and committed customer increases; If customers feel the guarantee of service quality and responsiveness of the employees of a company (Ree & Van, 2010).

IX. ASSURANCE

This result indicates the beta standardize is 0.319. Base on the result from Beta, the meaning of indicators that most consumers have a good response to the factor Assurance of Viettel Internet service provider. From there, customers can feel confident when contacting with Viettel services through the factor Assurance.

Recommendation

Based on Haron (1994), the employee's behavior, the cooperative, politeness, friendliness and the efficient customer handling is the most important factors for the customer satisfaction. A study by Leeds (1992), found that the customer satisfaction was increased by service quality and specialized conduct of company. Employee knowledge also has an impact on customer satisfaction because it gives the customer a sense of trust. Therefore, employees should contentiously polite in their customers, and the ability to answers the customer's questions are also well-mannered. Another study finds that when the company ensures that its employees are reliable, knowledgeable, and loyal; In addition, staff are very friendly and cooperative towards customers complaints and try to resolve them. These guarantees will create the trust of customers for the service of the company. (Madu, 2002)

X. EMPATHY

According to the multiple linear regression results, the relationship between empathy and customer satisfaction has been rejected. However, based on Munusamy et al (2010), Empathy is still important element of service quality so the company should sustain and improve the empathy. This can easily be seen through many policies of Viettel as offering many packages for customers to choose; for example, 6 packages for households and 4 packages for businesses or internet businesses.

REFERENCES

- [1] Aaker, D. (2004). Marketing research. Hoboken, NJ: John Wiley and Sons, Inc.
- [2] American Management Association. (2017). [online] Available at: <https://www.ama.org/> [Accessed 23 Sept. 2017].
- [3] Arasli, H., Smadi, S. M. & Katircioglu, S. T., 2005. Customer service quality in the Greek Cypriot banking industry. *Managing Service Quality*, 1(1), pp. 41-56.
- [4] ARNES, S., ENDERS, A. and JELASSI, T. (n.d.). E-commerce and V-business. Amsterdam: Elsevier.
- [5] Atkinson, K. and McDonald, A. (1992). Environmental management in Canada. Leeds: Regional Canadian Studies Centre, University of Leeds.
- [6] Balunywa, W. (1995). Business Administration, 1 st Edition. The business Publishinggroup Kampala Bowen and David E. (2005) Service Quality Blackwell encyclopaedic Dictionary of Human Resource management, P1 341, 340P
- [7] Brady, M. K. & Cronin, J. J. J., 2001. Some new thoughts on conceptualizing perceived service quality: a hierarchical approach. *Journal of Marketing*, 65(1), pp. 34-49.
- [8] Brady, M. K., Cronin, J. J. & Brand, R., 2002. Performance-only measurement of service quality: a replication and extension. *Journal of Business Research*, 55(1), pp. 17-31.
- [9] Brown, T.J., Churchill, G.A. and Peter, J.P. (1993) research note: Improving the measurement of service quality, *Journal of retailing*, 69(1) 126-139
- [10] Comrey, A. and Lee, H. (1992). A first course in factor analysis. New York: Psychology Press.
- [11] Comrey, A. L. & Lee, H. B. (1992). A first course in factor analysis. Hillsdale, NJ: Erlbaum.
- [12] Cronin, J. J., & Taylor, S. A. (1992) Measuring service quality; a re-examination and Extension. *The Journal of Marketing*, 56, (3), 55-68.
- [13] Culiberg, B. (n.d.). Consumers' right to be wrong.
- [14] Culture (2015) Oxford Dictionaries, view 23

- Sept, 2017. <http://www.oxforddictionaries.com/definition/english/culture>
- [15] Delgado-Ballester, E. (2004). Applicability of a brand trust scale across product categories: A multigroup invariance analysis, European Journal of Marketing 5 (6), 573-592.
- [16] Doanh. K, 2013. “ VNPT thong linh co dinh, Viettel dung dau dich vu di dong”. Pandora-Vietnam. Available at: <http://pandora.vn/vnpt-thong-linh-co-dinh-viettel-dung-dau-dich-vu-di-dong-0yXYaAMbxkpdE.html> [Accessed on 12th September, 2017]
- [17] Enis, B.M. (1974) “Marketing Principles: The Management Process”
- [18] Fornell, C. (1992). “A National Customer Satisfaction Barometer: The Swedish Experience.” Journal of Marketing, Vol. 56, p.6-21.
- [19] Gabbott, M. and Hogg, G. (1998). “Consumers and services”, Chichester: John Wiley & Sons.
- [20] Gamek, A 2017,”Cá 3 tuyến cáp quang biển của Việt Nam đồng loạt gặp sự cố, tốc độ Internet bị ảnh hưởng nghiêm trọng”, Gamek, [online] Available at: <http://gamek.vn/ca-3-tuyen-cap-quang-bien-cua-viet-nam-dong-loat-gap-su-co-to-cu-do-internet-bi-anh-huong-nghiem-trong-20170828150852082.chn>, [Accessed on 12 September. 2017].

Analyzing the Customer Satisfaction in Using Websites that Market Digital Products and Services

Dao Manh Cuong

FPT University
HCMC, Viet Nam

Nguyen Pham Bao Khanh

FPT University
HCMC, Viet Nam

Tran Cong Hau

FPT University
HCMC, Viet Nam

Do Hoang Tuan

FPT University
HCMC, Viet Nam

Abstract

In recent years, with the development of the Internet and high-tech services, E-Commerce is a potential and fast-growing market in Vietnam, according to reports of e-Marketer (2016), Vietnam E-Commerce Association (2015), and the E-Commerce and Information Technology (2014). In order to maximize the potential E-Commerce market of Vietnam, as well as understand the importance of studying the impact of Customer satisfaction in using websites that market digital products and services, the researchers decided to conduct the research. Because of the necessity of maintaining customer satisfaction, the researchers do this study to analyze the factor that affects to customer satisfaction in Vietnam electronic websites as well as in the context of MP3.ZING.VN & NHACCUATUI.COM. This research examines the application of theories in by developing a proposed research model. And then the authors use some research methods analyzing data to find out the practical meaning of Vietnam electronic commerce market.

We adopted a structured questionnaire-based survey comprising 26 questions and distributed to 150 responders within Ho Chi Minh City. The research focuses on the main customer from the result of Nielsen and Moore report in 2014. 225 questionnaires were returned and 150 questionnaires valued. On the other hand, the result of regression analysis proved three factors: Service quality, Information quality and Security have significant positive influence on customer satisfaction of Mp3.zing.vn and Nhacieuatui.com. In the other words, it is worthwhile for companies to use their resources by focusing on three vital factors for customer satisfaction.

Keywords

Customer Satisfaction

Supervisor: Vo Minh Hieu, MBA. (Head of Business Department, FPT University, HCMC FPT University)

Defended topic in front of the scientific council of HCMC FPT University on December 16th, 2017

Accepted post on December 26th, 2017

I. INTRODUCTION

The research articles will analyzing the customer's satisfaction in using three website are mp3.zing.vn, nhacieuatui.com.

In order to help assess the satisfaction of customers in using the digital product and service market in Vietnam, this research has several purposes as the following:

First, identify the factors that affect customer satisfaction in using website that market digital products and services.

Secondly, recognize the factors that make the highest and lowest satisfaction of in using website that market digital products and services.

Third, suggest some recommends improving customer satisfaction in using website that market digital products and services.

Based on these goals, the research find out what factors customers were not satisfied and what factors customers satisfied in using website that market digital products and services.

Moreover, this research also shows how different customers' satisfaction is when they use different website digital product and service. Different areas will have different solutions to improve customer satisfaction.

II. THEORETICAL BASIC AND LITTERATURE REVIEW

1. Theoretical basis and analysis framework

Definition of Digital good

Digital good means any software product or other good that is delivered or transferred electronically, including sounds, images, data, facts... where such good is the true object of the transaction, rather than the activity or service performed to create such good.

Definition of Customer Satisfaction

Customer Satisfaction is a measure of how happy customers feel when they do business with a company.

Definition of Buying Process

A buying process is a process that a consumer will take to make a purchasing decision.

Related Research: Wang, Tang and Tang .(2001); Gerbing and Anderson.(1988); Oliver and DeSarbo and Cote. (2000); Giese and Gote. (2000); Churchill

and Surprenant.(1982).

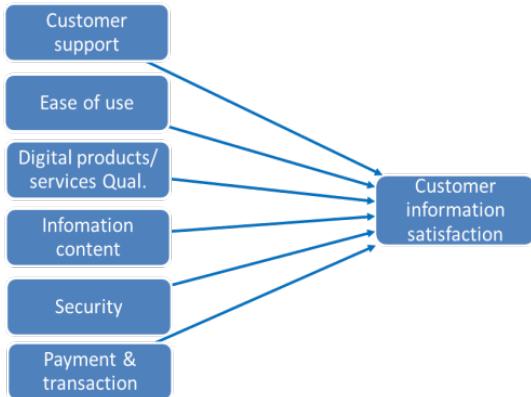


Figure 1

2. Literature review

Customer support

Customer support refers to customer services, feedback, and responses which can build the loyalty of customer and support for customer satisfaction.

- Wang, Tang, Tang, (2000) –

Ease of Use

The perceived ease of use is the customer perception that web page on the internet will involve minimum effort and it affect directly to customers satisfaction.

- Nadim and Noorjahan, (2007)-

Digital products/ services quality

Good quality will be the foundation for building customer satisfaction

- Oliver, (1993) -

Information Content

The easy way to approach to potential buyer is develop a web pages that can be displayed effectively information to e-user.

- S. Barnes and R. Vidgen. (2000) -

Security

Having a trusted website with exceptional security helps customers to be more confident and eventually, to feel satisfied.

- Gautam and Cynety, (2004)-

Payment and transaction

The evaluation of customer's satisfaction using online payment service helps test the relationships between service quality dimension customer satisfaction, customer trust and customer loyalty

- Wu et al, (2008) –

III. METHODS & DATA

1. Data

- Sampling methods: non-probability sampling method
- Research object: People from 15 – 45 years old who are using website that market digital products and services in HCMC
- Sample size: 465 samples included:

- 205 samples for mp3.zing.vn
- 204 samples for Nhaccuatui.vn
- 56 samples for pilot test
(Tabchinick & Fidell (1996) and Comrey and Lee (1992).)

- Data collection: Questionnaire survey

2. Methods/Framework

- Descriptive statistics: Descriptive statistics provides basic summaries of sample and scale in order to create a foundation for every data quantitative analysis

- Cronbach's Alpha: Cronbach's Alpha ratio is reliability ratio used to estimate correlation scale between observed variable-pairs

- Exploratory Factor Analysis (EFA): Exploratory Factor Analysis (EFA) is a statistical technique which is used to reduce observed variables to a smaller set but still maintains almost information content and statistic meaning of the initial set of variables (Hair et al., 1998).

- Regression Analysis: Regression analysis is one statistical method which is used to evaluate the strength of relations between one or more dependent factors and one or more independent factors.

IV. RESULT & FINDING

- Customer satisfaction in using website that market digital product and market was affected by the for 5 key factor include Customer Support, Ease of Use, Digital Product and Service, Information Content and Security.

- The finding does meet the initial expectation that accept 5 factors.

- Different when compare between two website and it suggest different methods for each model to modified and raise the customer satisfaction.

V. CONCLUSION AND POLICY

IMPLICATION

Customer Support:

- One of the biggest factor that have a significantly impact.

- MP3.Zing.vn is use C.Support to increase C.Satisfaction.

- Increase C.Satisfaction by changing C.Support.

- Need to increased C.Support MP3.Zing.vn 3.5 times, Nhaccuatui.com 5 times.

Ease of Use

- Using Mp3.zing.vn is easier than using Nhaccuatui.com.vn.

- Suggest for Nhaccuatui.com.vn should upgrade the EOU to raise C.Satisfaction.

Digital Products and Services.

- DP changes, the C.Satisfaction also changes.

- DP of MP3.zing.vn are better and more satisfying.

- Increase DP of MP3.zing.vn 5.5 times and Nhaccutui 6 times.
- Information Content.
- IC of Nhaccutui are better and more satisfying.
 - The good information content lead to good C.Satisfaction.
 - Increase IC of Mp3.zing.vn 5.5 times and Nhaccutui 5 times.

Security

- Increase Secure transactional data and privacy.
- Increase Security of Mp3.zing.vn 4 times and Nhaccutui 4.5 times.

Appendix (Heading 5)

REFERENCES

- [1] Brown, I. T. (2002). Individual and technological factors affecting perceived ease of use of web-based learning technologies in a developing country.
- [2] The Electronic Journal on Information Systems in Developing Countries, p.13 (3), 14-16.
- [3] C. Shannon, and W. Weaver, Mathematical: Theory of Communication, University of Illinois Press, Urbana, 1949
- [4] Chellappa, R.K. (2002). Customers' trust in electronic commerce transactions: The role of perceived privacy and perceived security. Emory University, Atlanta.
- [5] Comrey, A., & Lee, H. (1992). A first course in factor analysis. Hillsdale, NJ: Erlbaum.
- [6] Daniel E. Innis (1991), Customer service, repurchase intention, market orientation and firm performance in the channel, The Ohio State University.
- [7] Gautam and Cuneyt. (2004), Internet Shopping, Tong Zhou: Bang.
- [8] Giese, J. L. and J. A. Gote, defining Consumer Satisfaction Academy of Marketing Science Review [Online] Available: <http://www.amsreview.org/amsrev/theory/giese00-01.html>, 2000.
- [9] Guo, X., Ling, K.C., and Liu, M. (2012). Evaluating factors influencing customer satisfaction towards online shopping in China. Asian Social Science, 8(13), 40-50.
- [10] Karjaluoto, H., Mattila, M. and Pento, T. (2002), "Factors underlying attitude formation towards online banking in Finland", International Journal of Bank Marketing, ISSN 0265-2323, Vol. 20 No. 6, pp. 261-272.
- [11] Kennickell, Arthur B. and Myron L. Kwast (1997) Who Uses Electronic Banking? Results from the 1995 Survey of Consumer Finances," Working Paper, Board of Governors of the Federal Reserve System.
- [12] LaLonde, Bernard J. and Paul H. Zinszer (1976), Customer Services: Meaning and Measurement, Chicago: National Council of Physical Distribution.
- [13] Liao, Z., Cheung, M., 2002. Internet-based e-banking and consumer attitudes: an empirical study. *Information & Management* 39, 283–295.
- [14] Liu, S., Lin, Y., 2006. Grey Information Theory and practical Applications, Springer, London.
- [15] Mustafa, I.E. (2011). Determinants of e-commerce customer satisfaction, trust, and loyalty in Saudi Arabia. *Journal of Electronic Commerce Research*, 12(1), 78-93.
- [16] Nadim, J. & Noorjahan, B. (2007). The role of perceived usefulness and perceived ease of use. *African Journal of Business Management* p.32-40.
- [17] Nicosia, F. (1966) Consumer Decision Processes, Prentice Hall, Englewood Cliffs, 1966.
- [18] Rencher, Alvin C.; Christensen, William F. (2012), "Chapter 10, Multivariate regression – Section 10.1, Introduction", Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, 709 (3rd ed.), John Wiley & Sons, p. 19,
- [19] Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd ed.). New York: HarperCollins.
- [20] Takieddine, S., Sun, J., 2015. Internet banking diffusion: A country-level analysis. *Electronic Commerce Research and Applications*. Vladimirov, Z., 2012. Customer satisfaction with the Bulgarian tour operators and tour agencies' websites. *Tourism Management Perspectives* 4, 176–184.
- [21] W.H. DeLone, and E.R. McLean, "Information Systems Success: the Quest for the Dependent Variable", *Information Systems Research*, 3(1), 1992, pp. 60 - 95. Management Dynamics in the Knowledge Economy, 2017, Vol5.
- [22] Wu, YL., Chang, M., Yang, CH., Chen, YJ., 2008. The Use of E-SQ to Establish the Internet Bank Service Quality Table. *Proceedings of the IEEE IEEM*.
- [23] Yoon, CH., 2010. Antecedents of customer satisfaction with online banking in China: The effects of experience. *Computers in Human Behavior* 26, 1296–1304.
- [24] Zeithaml, V., Parasuraman, A., Malhotra, A., 1999. Service Quality Delivery through Web Sites: A Critical Review of Extant Knowledge. *Journal of the Academy of Marketing Science* 4, 362-375.
- [25] Farrar, Donald E.; Glauber, Robert R. (1967). "Multicollinearity in Regression Analysis: The Problem Revisited". *Review of Economics and Statistics*.

Analyzing the Influence Factors to Consumer Acceptance of Internet of Things Technology

Pham Quoc Trung

FPT University
HCMC, Viet Nam

Nguyen Thanh Hoai Ngan

FPT University
HCMC, Viet Nam

Ho Thi Dieu Huyen

FPT University
HCMC, Viet Nam

Phung Mai Khanh

FPT University
HCMC, Viet Nam

Abstract

In recent years, *Internet of things technology* (abbreviated as *IoTs*) is a term which appeared quite a lot and attracted a lot of attention of the tech world and is expected to explode in the future. It will have a strong impact on the life, working, and society, including Vietnam. To maximize the potential of Vietnam's Internet market, especially in Ho Chi Minh city, researchers have decided to conduct research. The objective of this study is to identify factors that impact directly and indirectly to consumer acceptance; to recognize the factors affecting the acceptance of consumers from high to low; to identify and define demographic difference for the Internet of things technology.

The research method includes both qualitative research (in-depth interview and focus group) and quantitative research (survey). The valid sample size of the survey is 352. The research model is assessed by using the procedure of referring a number of previous studies and scholars, analyzed by SPSS and the most influential factors. Within this research, we want to find the influence factor to consumer acceptance of Internet of thing technology, and we also figure out how they affect this intention. The result shows that there are six main factors which positively affect consumer acceptance about the Internet of things technology. They have perceived usefulness, ease of use, trust, social influence, enjoyment and behavioral control.

Keywords

Consumer Acceptance

I. INTRODUCTION

The term internet of things (*IoTs*) first appeared in 1999 by Ashton (2009), a British technologist.

He was also the first to help develop this concept (Gubbi et al., 2013). *IoTs* extends the benefits of Internet connectivity, remote control, and data sharing, etc. (Peoples et al., 2013). At the most basic level, *IoTs* involves the assembly of microelectronic objects and communication antennas (Deutsche Welle, 2012). Using radio frequency identification (RFID), all real objects in the analog world can have a unique identity, such as IP address (Gubbi et al., 2013).

In Vietnam, this concept has also become a hot topic, *IoTs* also is talked the most about all IT events in the IT industry in 2016 (Ngoc Bich, 2017). The technology experts said that all object will integrate software, sensors to connect with each other and interact with people to create a world of internet connection in the future, which will change the people live, making our life more convenient for smart connections (Ngoc Bich, 2017).

Commenting in future technology trends, FPT Board Chairman Truong Gia Binh said that: Internet of things technology is a potential technology trend, Vietnam cannot stand outside the global game (Ngoc Bich, 2017). Currently, FPT is researching a number of solution to connect to the Internet as smart city; intelligent traffic (basic management systems such as traffic signal control systems, container management systems, message boards, automatic number plate recognition, etc.), home appliance control center (smart home), building a platform providing solutions services for *IoTs*... Especially, FPT has deployed projects connected in the field of healthcare. Internet connectivity is expanding the reach, not just between people, but also with devices, processes, data and objects (Ngoc Bich, 2017). Cisco, a provider of IT solutions and equipment, also said that in 1984 when Cisco was founded only had about 1000 devices were connected to the World Wide Web, by 2010, this figure up to 10 billion (Ngoc Bich, 2017).

II. THEORETICAL BASIC AND LITTERATURE REVIEW

Theoretical basis and analysis framework

This study will consider emotion and perceptual

Supervisor: Vo Minh Hieu, MBA. (Head of Business Department, FPT University, HCMC FPT University)

Defended topic in front of the scientific council of HCMC FPT University on December 16th, 2017

Accepted post on December 26th, 2017

orientation of an individual to determine consumer acceptance. Constructing intention to use is passed and seen as demonstrate of consumer acceptance (Mathieson, 1991; Venkatesh and Davis, 2000). The model describing the relationship between variables and behavioral intentions to use IoTs technologies is shown in Figure 1. Theoretically, many studies have shown a clear and strong causal between the user's trust and the behavioral intent to use the technology (Lee et al., 2012). Therefore, the above theory is expressed through behavior intention to use and it is considered as a dependent variable to examine the acceptance of IoTs technology. Research includes the perceived usefulness factors, easy to use perceptions, trust, social influences, and perceived enjoyment will have impact on behavioral intention to use of IoTs technology.

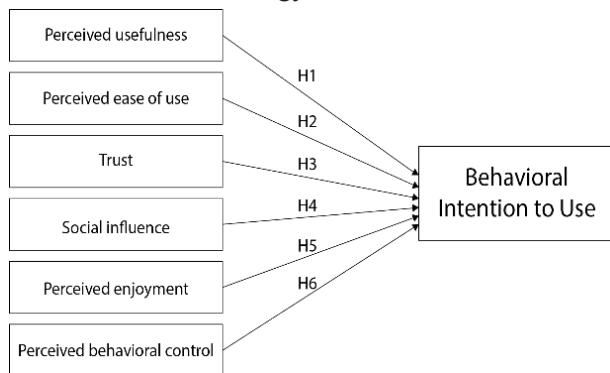


Figure 1. Consumer acceptance of IoTs

III. METHODS & DATA

1. Data

- Sampling methods: non-probability sampling method
- Research object: People from 15 – 45 years old who are using website that market digital products and services in HCMC
- Sample size: 400 samples included:
 - 150 online samples
 - 350 offline samples
 - 50 samples for pilot test
 - 352 sample : accepted

(Tabchinick & Fidell (1996) and Comrey and Lee (1992).)

- Data collection: Questionnaire survey
- Location Scope

In Ho Chi Minh City, the Internet is very popular, so we made a lot of places surveyed many different ages, different professions to provide the most accurate results. Therefore, research is done in many places: Lac Viet Co., Eservice Group Co., The Local coffee, FPT University and International Neurosurgery hospital via direct survey.

- Time Scope

According to time scope, this research collects the data from September, 16th 2017 to November, 3rd 2017.

- Age scope

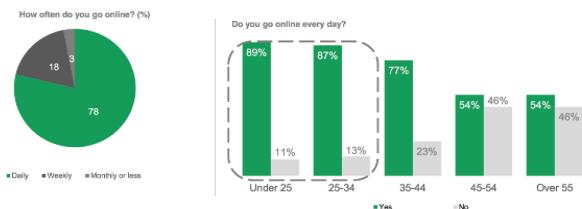


Figure 2

According to research about the connected consumer in Vietnam following Consumer Barometer report 2015, we can see that especially consumers under 34 years old often use the internet. Thus, the research will concentrate all of the consumers from under 25 to 34 years old.

2. Methods/Framework

Acceptance of using of technology services:

It means that the consumer has tried the product and accepted the inquiry, collected the opinion, enhanced the usability to continue using the service. The act of using is not only compulsory or externally imposed but also is based on voluntary willingness to use the service for its own purposes

- Demographic questions are designed to help researchers determine what factors affect to behavioral intention to use. In this part, we only have two questions involve age and gender:
 - How old are you?
 - What is your gender?
- Cronbach's Alpha: Cronbach's Alpha ratio is coefficient between the variables-total will help to exclude those variables that do not contribute much to the description of the concept under investigation
- Correlation analysis : Pearson correlation coefficient (denoted r) is used to measure the correlation between independent variables and dependent variables. In principle, Pearson's correlation will find the line that best matches the relationship between two variables
- Exploratory Factor Analysis (EFA): The method is used in many fields such as evaluate and correlation among variables or interpretive questionnaire (Stevens, JP 1992). Besides, EFA also is used to check the level, the structure, the influence of variables on the scale and can grouping for variables.
- Regression Analysis: Regression analysis used to determine the specific index of each independent factor affecting the dependent factor. By using an equation of a straight line, it can analyze

the linear relationship between dependent variable and independent variables. This is the general formula:

$$Y = \text{Constant} + B^*X_n$$

IV. RESULT & FINDING

The survey was conducted via directly and online method to the respondent with 352 questionnaires. Based on the characteristic trait of the object such as ages and places, the survey was conducted directly at International Neurosurgery Hospital (Tan Phu District), FPT University (District 12), The local coffee (District 1). Besides, some results of the questionnaire are not valid. Because, many people do not answer surveys honestly; for example, they just fill the answer with one level from start to end such as all of 1 or 5.

V. CONCLUSION AND POLICY IMPLICATION

In summary, the research objectives have been achieved. The key factors which affect the research object and the relationships between them are clarified. The research gives overviews about evaluating the importance of critical factors to customer acceptance of Internet of Things Technology. Although all of the objectives of the study have been achieved, the following the process of exploring from previously available sources to our research, data are collected and analyzed by SPSS, we think there will be many work done in the future to improve the limitations. Overall, with the research proposal, there are six factors that affect the acceptance of the customers: Perceived usefulness, perceived ease of use, trust, social influence, perceived enjoyment, perceived behavioral control. According to this study, all factors have a significant positive impact on customer acceptance, in other words, the increased value of those factors will lead to increased motivation, Internet usage of the customer's behavior, and many manufacturers will improve the product to fit the needs of the majority of customers, improve customer satisfaction in the future

REFERENCES

- [1] Ajzen and Fishbein's Theory of Reasoned Action (TRA) (1980), Available from: safaribooksonline E-Book, 50-55
- [2] Ajzen, I. (1985), From intentions to actions: A theory of planned behavior, Springer, New York.
- [3] Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211 at <https://pdfs.semanticscholar.org/6256/ca4853f44ab9acb98f91f0d7848c54185ca7.pdf> [Accessed 15 October 2017]
- [4] Daniel Boduszek "Standard Multiple Regression", University of Huddersfield, UK, pp11-13
- [5] Davis, F.D (1989), "Perceived usefulness, perceived ease of use and user acceptance of information technology", MIS Quarterly, 13(3), 319-339
- [6] Davis, F.D(1993)," User acceptance of computer technology: System characteristics user perceptions and behavior characteristics", International Man-Machine studies, 38, 475-487
- [7] Denzin, Norman K. and Lincoln, Yvonne S. (2000, 2ed ed.). *Handbook of Qualitative Research*. London: Sage. <http://atlasti.com/qualitative-research/> [Accessed 23 October 2017]
- [8] Deutsche Welle (2012), Internet of Things' Holds Promise, But Sparks Privacy Concerns, available at: www.dw.de/internet-of-things-holds-promise-but-sparks-privacy-concerns/a-15911207-1 [Accessed 23 October 2017]
- [9] Ding, W. (2013), "Study of smart warehouse management system based on the IOTS", in Du, Z (Ed.), *Intelligence Computation and Evolutionary Computation*, Springer, Berlin,pp. 203-207.
- [10] Doolin, B., Dillon, S., Thompson, F. and Corner, J.L. (2005), "Perceived risk, the internet shopping experience and online purchasing behavior: a New Zealand perspective", *Journal of Global Information Management*, Vol. 13 No. 2, pp. 66-88.
- [11] Gubbi, J., Buyya, R., Marusic, S. and Palaniswami, M. (2013), "Internet of things (IoTs): a vision, architectural elements, and future directions", *Future Generation Computer Systems*, Vol. 29 No. 7, pp. 1645-1660
- [12] Meyers, L., Gamst, G. and Guarino, A. (2013). *Performing data analysis using IBM SPSS®*. New Jersey: John Wiley & Sons, Inc.
- [13] Nielsen company (24th November, 2017) , "Smartphone keep up the growth momentum with increasing usage incidences in key city", http://www.nielsen.com/content/dam/nielsenglobal/vn/docs/PR_EN/Web_Nielsen_Smartphones%20Insights_EN.pdf [Accessed 20 October 2017]

Conference Participants

Huynh Cong Viet Ngu, M.E

Lecturer of IT Department, HCMC FPT University
nguhcv@fe.edu.vn

Nguyen Anh Luan

Student at HCMC FPT University
t14.luanna@gmail.com

Tran Phuoc Tien

Student at HCMC FPT University
neoandesonn@gmail.com

Mai Duc Thang

Student at HCMC FPT University
maiducthang97@gmail.com

**Students of Sofware Engineering &
Information Assurance Faculty,
HCMC FPT University**

Vuong Chanh Quyen

quyenvcse62765@fpt.edu.vn

Trinh Thanh Quyen

quyenttse62766@fpt.edu.vn

Nguyen Minh Nhut

nhutnmse62761@fpt.edu.vn

Nguyen Viet Cong Hau

haunvcse61546@fpt.edu.vn

Luu Nguyen Ngoc Thu

thulnnse61997@fpt.edu.vn

Nguyen Ngo Duc Khai

khainndse60889@fpt.edu.vn

Le Vuong Quoc Huy

huylvqse61844@fpt.edu.vn

Nguyen Chi Hieu

hieuncse61483@fpt.edu.vn

Pham Bao Tin

tinpbse61681@fpt.edu.vn

Do Viet Son

sondvse61439@fpt.edu.vn

Ngo Quang Tuan Anh

anhnqtse61456@fpt.edu.vn

Vo Thanh Danh

danhvtse61518@fpt.edu.vn

Doan Vu Tung Lam

lamdvtse60790@fpt.edu.vn

Nguyen Thanh Tung

tungntse61413@fpt.edu.vn

Huynh Huu Nhan

nhanhhse61453@fpt.edu.vn

Nguyen Thi Phuong Nhi

nhintpse61552@fpt.edu.vn

Tran Nguyen Bao Duy

duytnbse61589@fpt.edu.vn

Ma Lam Ngoc

ngocmlse62083@fpt.edu.vn

To Thanh Thuyen

thuyenttse61437@fpt.edu.vn

Nguyen Hoang Tuan

tuannhse61787@fpt.edu.vn

Nguyen Quang Huy

huynqse61525@fpt.edu.vn

Nguyen Tuan Sang

sangntse61499@fpt.edu.vn

Mai Thanh Hieu

hieumtse61559@fpt.edu.vn

Huynh Quang Dat

dathqse61783@fpt.edu.vn

Tran Hong Son

sonthse61470@fpt.edu.vn

Le Xuan Minh

minhlxse61839@fpt.edu.vn

Vo Thong Minh

minhvtse61564@fpt.edu.vn

Tran Quoc Phuong Nam

namtqpse61829@fpt.edu.vn

Do Duy Thinh

thinhddse61815@fpt.edu.vn

Nguyen Xuan Huy

huynxse60857@fpt.edu.vn

Bui Cong Khanh

khanhbcse60847@fpt.edu.vn

Bui Quang Vien

vienbqse61466@fpt.edu.vn

Nguyen Khai Tri

trinkse61538@fpt.edu.vn

Nguyen Hoang Phuc

phucnhse61520@fpt.edu.vn

Nguyen Thi Thai Han

hanntse61600@fpt.edu.vn

Nguyen Van Bach

bachnvse61455@fpt.edu.vn

Phan Nhat Minh
minhpnse61705@fpt.edu.vn

Mai Minh Quy
quymmse61610@fpt.edu.vn

Truong Huu Thanh
thanhhse61493@fpt.edu.vn

Phan Minh Huan
huanpmse61860@fpt.edu.vn

Pham Trung Hieu
hieuptse60874@fpt.edu.vn

Ho Bao Tien
tienhbse61862@fpt.edu.vn

Hoang Quoc Hung
hunghqse62049@fpt.edu.vn

Lam Ngo Quyen
quyenlnse61838@fpt.edu.vn

Huynh Thanh Trung
trunghtse61581@fpt.edu.vn

Tuong Duc Khang
khangtdse61752@fpt.edu.vn

Nguyen Van Hung
hungnvse61305@fpt.edu.vn

Duong Hai Quang
quangdhse61383@fpt.edu.vn

Thai Hoang Duy Bao
baothdse61666@fpt.edu.vn

Hoang Vinh Son
sonhvse61570@fpt.edu.vn

Nguyen Dinh Thong
thongndse61656@fpt.edu.vn

Vu Thai Nam
namvtse61522@fpt.edu.vn

Huynh Ngoc Dat
dathnse61688@fpt.edu.vn

Ngo Nguyen Thuy Van
vannntSE61626@fpt.edu.vn

Le Thanh Danh
danhltsse61111@fpt.edu.vn

Vo Manh Hung
hungvmse61352@fpt.edu.vn

Nguyen Le Minh
minhnlse61530@fpt.edu.vn

Vo Phuc Hai

haivpse61582@fpt.edu.vn

Ton That Minh Tri
trittmse61679@fpt.edu.vn

Nguyen Phu Ngoc Trai
trainpnse61060@fpt.edu.vn

Nguyen Kim Cuong
cuongnkse90180@fpt.edu.vn

Students of Business Management Faculty, HCMC FPT University

Vo Thi Tuyet Nhi
nhuvtsb60861@fpt.edu.vn

Le Thi Diem Xuan
xuanltdsb60803@fpt.edu.vn

Vo Quang Sang
sangvqsb60705@fpt.edu.vn

Pham Anh Duc
ducpasb60678@fpt.edu.vn

Dao Manh Cuong
cuongdmsb60881@fpt.edu.vn

Nguyen Pham Bao Khanh
khanhnpbsb60584@fpt.edu.vn

Tran Cong Hau
hautcsb60489@fpt.edu.vn

Do Hoang Tuan
tuandhsb60661@fpt.edu.vn

Pham Quoc Trung
trungpqsb60812@fpt.edu.vn

Nguyen Thanh Hoai Ngan
ngannthsb60977@fpt.edu.vn

Ho Thi Dieu Huyen
huyenhtdsb60846@fpt.edu.vn

Phung Mai Khanh
khanhpmsb60830@fpt.edu.vn

Awards & Photos Appendix

Evaluation criteria	
► The score of the review committee accounted for 60%	
► Audience at the auditorium vote account for 40%: Audience at the auditorium vote according to the criteria: Application of the subject in practice; Presentation ability of the group.	
Total score: 100 points	
Level	
Very good (100 points)	
Good (90 points)	
Normal (80 points)	
Low (70 points)	
Conference Awards	
First Prize:	
Hacking and Security Techniques on Web Applications	
Ho Bao Tien, Lam Ngo Quyen, Huynh Thanh Trung, Hoang Quoc Hung	
Second Prize:	
University Admission Counseling System for High School Students	
Ngo Nguyen Thuy Van, Le Thanh Danh, Vo Manh Hung, Nguyen Le Minh	
Third Prize:	
Applying iBeacon Technology to build Automation Tollbooth Station	
Le Vuong Quoc Huy, Do Viet Son, Pham Bao Tin, Nguyen Chi Hieu	
Consolation Prize:	
1. Analyzing the Consumer Satisfaction for an Internet Service Provider	
Vo Thi Tuyet Nhi, Le Thi Diem Xuan, Vo Quang Sang, Pham Anh Duc	
2. Analyzing the Influence Factors to Consumer Acceptance of Internet of Things Technology	
Pham Quoc Trung, Nguyen Thanh Hoai Ngan, Ho Thi Dieu Huyen, Phung Mai Khanh	
3. FootBall Field Reservation System	
Mai Minh Quy, Truong Huu Thanh, Phan Minh Huan, Pham Trung Hieu	
4. Warning System on Parameters of the Devices in the Family	
Vo Phuc Hai, Ton That Minh Tri, Nguyen Phu Ngoc Trai, Nguyen Kim Cuong	

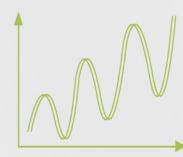
ABSTRACT



Time series data play an important role in various fields from science, technology, medicine to economics such as user data management, business planning, meteorological forecasting, etc. Therefore, clustering time series data is essential to classify time series and develop Decision Support Systems (DSS). So, a framework, running on Windows, is designed and implemented for clustering time series data.

INTRODUCTION

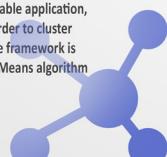
In our life, time series is a common type of dynamic data that naturally arises in many different scenarios, such as economic data, forecast data, electrocardiogram, etc. Clustering time series data is make them into clusters in order to having an overview of the differences of each cluster as well as the nature of them.





PROBLEM & SOLUTION

Nowadays, with the incremental development of data storage and retrieval, data clustering, especially time series, has been shown effective in providing useful information in various domains. Because of the needs of clustering data and reusable application, the framework must be created. In order to cluster time series quickly and efficiently, the framework is designed and developed based on K-Means algorithm and square of Euclidean distance.



CONCLUSION

Advantages

- The algorithms implemented in the framework have high accuracy.
- Data visualization is supported to help users cluster new and unclassified data.

Disadvantages

- Data normalization is not supported. So users must spend time normalizing data into a uniform.
- The framework is not possible to cluster time series data having high noise and complexity.



Subject's Poster: A Framework for Clustering Time Series Data

Apply Collaborative Filtering to Give Suggestions for Users in Web Application

Nguyen Viet Cong Hau, Luu Nguyen Ngoc Thu, Nguyen Ngoc Duc Khai
FPT University, Ho Chi Minh City, Vietnam

Abstract

Nowadays, websites are one of the most important things in human life. People access and go online every day, every hour, every hour and every second. Having too much information in the "world" - world of websites is too much. Users have many choices to view. How do users select suitable items to view, in dozens, hundreds or thousands (or more) items in a website? How do website owners can keep user staying with them if they do not give "the right item" for "the right people"? That is why websites need to have some ways to give recommended items to their users. These items should be something that users like.

Keywords: recommendation, suggestion, collaborative filtering, pearson correlation score, recommended item

Introduction

The basic method that a website may use to display its items to users is category filtering. It seems good because now the user can buy a new product. However, user needs do many steps to get the right product. This method may be good if user really know what they want. What if user did not really know what the exactly product they would like to buy? What if user just "want to buy something"?

We assume that a user knows what kind of product they would like to buy, for example, "a cellphone". If you were the owner, which product would you like to sell to this customer? A "feature phone" or a smart phone? A "cheap smart phone" or a "flagship" smartphone?

It seems that there are product categories that has most searching times from the other users, or just simply categories which's products is "bestelling" products. Is recommending "bestelling" products or top search products an effective way to make customer buy more?

Solution

We would like to introduce another effective method to give suitable recommended items to website's users. It is a method introduced by Toby Segaran in the book "Programming Collective Intelligence" published on August 2007 by O'Reilly to solve this problem. It is the method Collaborative Filtering.

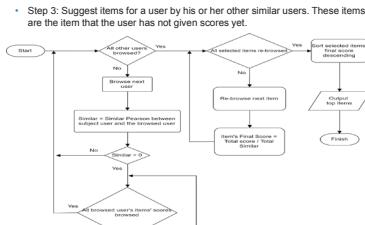
It is the method that give recommended items to a user base on the data from the other users who are similar to the user.

That method is finding the other users that have "same taste" with the user. When two people has same taste or similar together, we may use this person's experience to give recommendations to the other.

Plan Implementation

With Collaborative Filtering, we need to do these following steps to suggest items to an user:

- Step 1: Build a dataset that is the feedback score users give to items.
- Step 2: Find similar users by look for users that has similar feedback to a common (between 2 users) item set.
- Step 3: Suggest items for a user by his or her or similar users. These items are the item that the user has not given scores yet.



Experiment Result

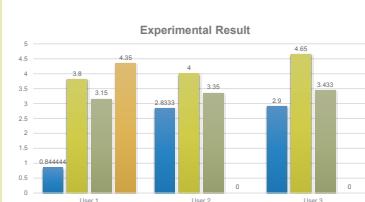
To understand more this method, we should use the example below. The target is now give suggested experts by the collaborative filtering. Assume that we have a dataset of users' feedbacks' scores to expert each time they receive a help.

Assume that the user need to be suggested experts is user 4. Then we use the scores above to calculate the pearson correlation score. Then, we could get the similar users to the user 4 as the list below.

Finally, we list out the scores of all other users give to experts that user 4 has not given scores yet

User	Product	Score
1	1	0.044444
1	2	3.8
1	3	3.15
1	5	4.26
1	8	3.43333
2	1	2.03333
2	2	4
2	3	3.35
2	9	3.9
3	1	2.9

User	Product	Score
5	3	4.05
6	3	4.7
6	12	1.2
6	44	4.45
7	43	2.8
7	44	3.9



Experiment Results

User	Product	Score
User 1	Expert 1	0.044444
User 1	Expert 2	3.8
User 1	Expert 3	3.15
User 1	Expert 4	4.26
User 1	Expert 5	3.43333
User 2	Expert 1	2.03333
User 2	Expert 2	4
User 2	Expert 3	3.35
User 2	Expert 4	4
User 2	Expert 5	3.9
User 3	Expert 1	2.9
User 3	Expert 2	4.05
User 3	Expert 3	4.7
User 3	Expert 4	1.2
User 3	Expert 5	4.45
User 4	Expert 1	3.43333
User 4	Expert 2	4.26
User 4	Expert 3	3.8
User 4	Expert 4	4.45
User 4	Expert 5	3.9

Conclusions

From the above result, we could list out the suggested products for user 4 are: expert 2, expert 1.

References

Programming Collective Intelligence
By Toby Segaran, published on August 2007 by O'Reilly.
Copyright © 2007 Toby Segaran. All rights reserved
ISBN-10: 0-596-52932.ISBN-13: 978-0-596-52932-1

Subject's Poster: Apply Collaborative Filtering to Give Suggestions for Users in Web Application

116

Posters

Customer Relationship Management for QTSC

INTRODUCTION

At present, QTSC in particular and others company in general is managing its sales relationship with customer through a standard method such as by document, verbal communication, and some others direct communication.

SOLUTION

- Devide system user into several roles
- Don't give an export function
- Design an intuitive interface

ADVANTAGE

- Suitable for QTSC procedures
- Resolve wasting-time in managing
- Resolve duplicating customer data
- Easy to use

PROBLEM

- Facing to be leaked customer data because of untightly data management.
- Work performance in the company is extremely low by manage data by hand.
- Reporting is always take too much time on current process.

ANALYSIS

- We choose .Net Web API 2 to build this system because it's a great framework and also is a open source.
- And we do create an Filter input data from customers algorithm to help this system more powerful.

FUTURE PLAN

- Be able to connect to other applications
- More secure
- Less server scrashing

Subject's Poster: Customer Relationship Management System for QTSC




Electronic Devices B2B Website

Instructor: Mr. Lại Đức Hùng
Member: Nguyễn Hoàng Phúc
Nguyễn Thị Thái Hân
Phan Nhật Minh
Nguyễn Văn Bách

1 Online ordering

The website allow users to search for products based on categories, brands. Then the users can order products via this system. Finally, the products will be delivered to the given addresses by suppliers. The website plays an intermediary role in order to ensure secure and smooth transactions.

2 Online negotiation

The website supports users in real-time interacting with supplier and through an online messaging system aiming to maximize the success possibilities of transactions. When a buyer has special requirements for an order, he or she can contact the supplier directly through this function.

3 Creating bidding contracts

The website assists users in seeking for trustworthy suppliers through a function which creates bids and assesses contractors. The supplier will provide the best criteria for winning the bid. Consequently, users can make better assessment and build trust among prospective partners.

4 Group-buying

The website allows users (suppliers) create group-buying deals so that customers can get discounts. Customers can buy products at a low price. Besides, suppliers can also make more sales as well as expand the sales market through this measure.

REFERENCES

- [1]https://www.gocom.vn/
- [2]https://www.alibaba.com/
- [3]https://machinelearningcoban.com/2017/05/24/collaborative-filtering/
- [4]http://socketo.me/docs/
- [5]http://blog.rikkeisoft.com/seminar-gioi-thieu-ve-websocket-va-node-js/
- [6]https://sparktutorials.github.io/2015/11/08/spark-websocket-chat.html

Subject's Poster: Electronic Devices B2B Website

UNISTART

Introduction

Students are really need informations about university's introduce, major, score and how other peoples, specially is the students of that university think about it. Student nowadays can't find the fully and constructed information about all university to make a choice. Some university's information gates like kenhtuyensinh.vn, tuoitre.vn....

All informations are put in article format so that it is difficult for students to search the informations that they want. And the most importance, other information systems don't have personality test to help students know which career is suitable for them, give them an recommendation for suitable university.

Instead of a general information, we will focus more on searching, rating, review and recommendation by user's information. The system also provide the personality test to help student find what they want.

Suitability Recommendation:

This recommendation is the way to suggest university for user. This recommendation based on result MBTI of user and review about university.
Suitability = $0,7 * a + 0,3 * b$

a: number of similar majors of university with suggested major based on MBTI type
b: The largest number of similar majors in list S
c: recommended study university point (0 to 1)

University admission counseling system for High School Student

Instructor: Lam Huu Khanh Phuong
Ngo Nguyen Thuy Van - Le Thanh Danh - Vo Manh Hung - Nguyen Le Minh

Correlation Recommendation

$$C = \left[\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} + 1 \right) * \frac{n}{t} \right] * 0,6 + a * 0,2 + b * 0,2$$

n: Number of similar majors.
t: The largest number of similar majors.
x: The origin university's latest year score
y: The university's latest year score needs correlation
a: Location point (0 or 1)
b: Training system point (0 or 1)

References

- <https://kenhtuyensinh.vn>
- <https://www.tuoitre.vn>
- <https://diemthituyensinh247.com>
- <https://maximusveritas.com>

Subject's Poster: University Admission Counseling System for High School Students

STARS RATED FOR EXCELLENCE 2012

FPT Education

FPT UNIVERSITY

Automation Tollbooth System

Lê Vuong Quốc Huy
Nguyễn Chí Hiếu
Phạm Bảo Tín
Đỗ Việt Sơn

Supervisor: Kiều Trọng Khanh

Subject's Poster: Applying iBeacon Technology to build Automation Tollbooth Station

STARS RATED FOR EXCELLENCE 2012

FPT Education

FPT UNIVERSITY

C M E

Building CROWD MONITORING OF EXPO System using iBeacon Technology

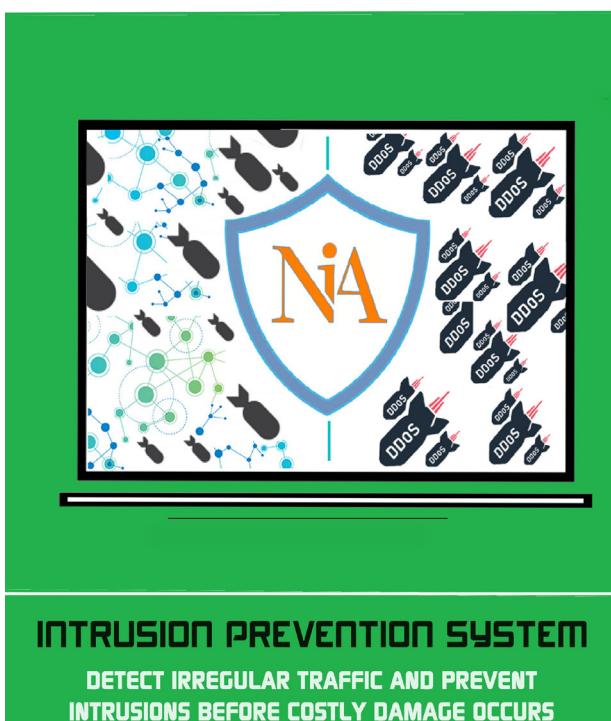
When an exposition is held, there are a number of stalls with a lot of visitors and some with few visitors. The exhibition organizers often count the attractiveness of each stall by eye at a specific time that the organizer decides. This statistic is made subjective at a specific time that the organizer decides. This makes the statistics and analysis of user trends in the exhibition is inaccurate. Beside that, they can be made by survey with visitors, or counting the number of given gifts. However, the exhibition organizers can't force all visitors do survey and control the delivery gifts equally.

In the process of analysis, we believe the beacons technology is capable to complete goals that we set out with the ability to build a device grid that can indicate visitors' trends at exhibition.

The system includes the numbers of beacons, some mobile devices and a web application. We use beacon combined with mobile application to collect the number of the visitors in the exhibition at any time. From these figures, we can calculate and analysis to generate the heat map and report for expo organizer.

Supervisor	Kiều Trọng Khanh
Member	Võ Thành Danh
	Ngô Quang Tuấn Anh
	Đoan Vũ Tùng Lâm

Subject's Poster: Building Crowd Monitoring of Expo System Using iBeacon Technology



Subject's Poster: DDoS Attack Prevention System



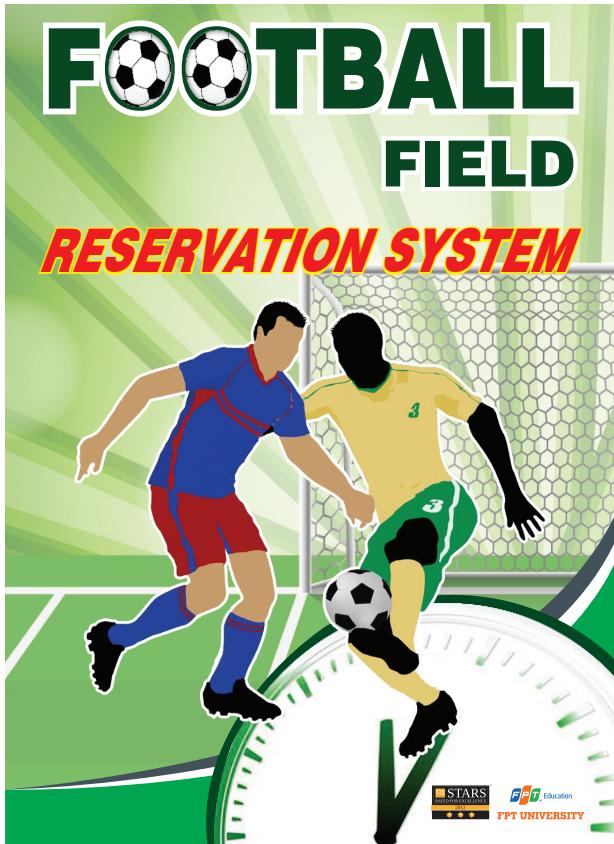
Subject's Poster: Develop a Security Information Event Management System



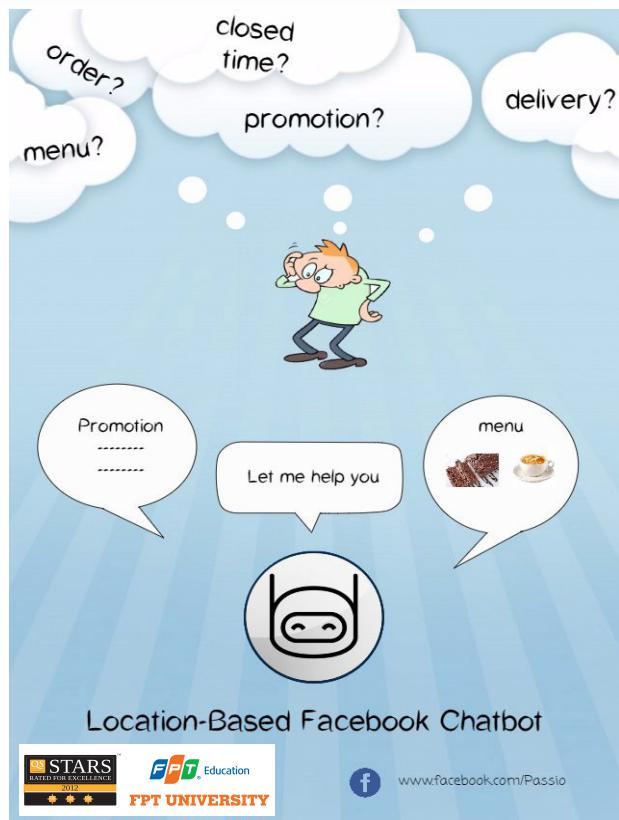
Subject's Poster: Dendrobium Orchid Care Solution



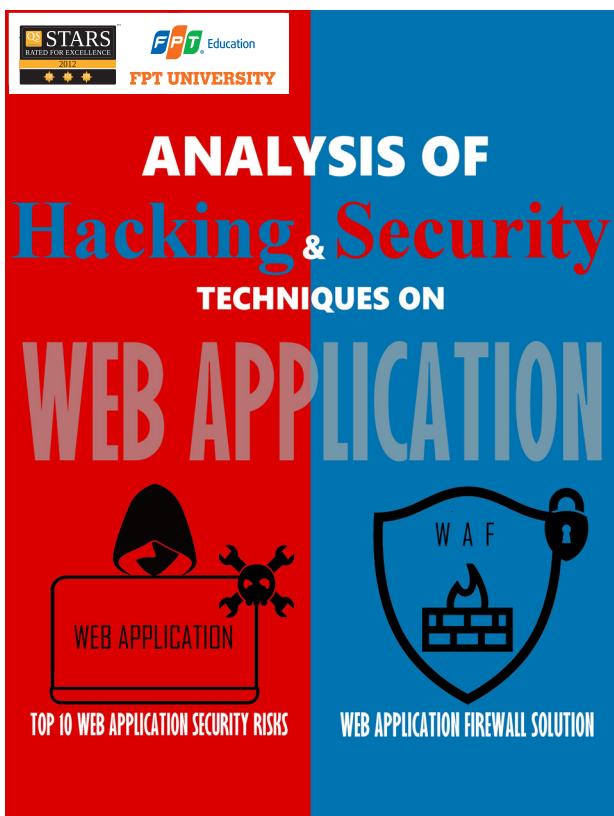
Subject's Poster: Door Access and Payroll Monitoring System



Subject's Poster: Football Field Reservation System



Subject's Poster: Intent Analyzer, Dialog Flow Controller and Sentence Suggestion



Subject's Poster: Hacking and Security Techniques on Web Application



Subject's Poster: sProcurement



of devices in the family

For the people at Any Love IOT Technology

[CONTACT US
FOR MORE INFORMATION](#)

**Subject's Poster: Warning System on
Parameters of Devices in the Family**

The poster has a dark blue background with white and orange icons related to technology and consumer behavior. Logos for STARS and FPT University are at the top. The main title is 'Analyzing the influence factors to CONSUMER ACCEPTANCE of Internet of Things Technology'. It includes author information (Ho Chi Minh, September 2017), member names (Pham Quoc Trung, Nguyen Thanh Hoai Ngan, Ho Thi Dieu Huyen, Phung Mai Khanh), and supervisor information (Vo Minh Hieu). A yellow circle on the right says 'Group 4'. A detailed description of the research method is provided at the bottom.

In recent years, Internet of things technology (abbreviated as IoT) is a term which appeared quite a lot and attracted a lot of attention of the tech world and is expected to explode in the future. It will have a strong impact on the life, working, and society, including Vietnam. To maximize the potential of Vietnam's Internet market, especially in Ho Chi Minh city, researchers have decided to conduct research. The objective of this study is to identify factors that impact directly and indirectly to consumer acceptance; to recognize the factors affecting the acceptance of consumers from high to low; to identify and define demographic difference for the Internet of things technology.

Subject's Poster: Analyzing the Influence Factors to Consumer Acceptance of Internet of Things Technology



Subject's Poster: Analyzing the Consumer Satisfaction for an Internet Service Provider



Mr. Than Văn Su (Head of Academic Affairs) was speaking to start the Conference



The Scientific Council was inspecting presentation of teams



Mr. Huynh Công Việt Ngu was presenting Keynote Speech 1: “How to write a Research paper for International Publishing”



The Students was inspecting presentation of teams



FPTU's Students was presenting Keynote Speech 2: “Pollia: A New Approach of Web Application for Tourism Service”



Team 1 was presenting subject: “Analyzing the Consumer Satisfaction for an Internet Service Provider”



Team 2 was presenting subject: “Analyzing the Influence factors to Consumer acceptance of Internet of things Technology”



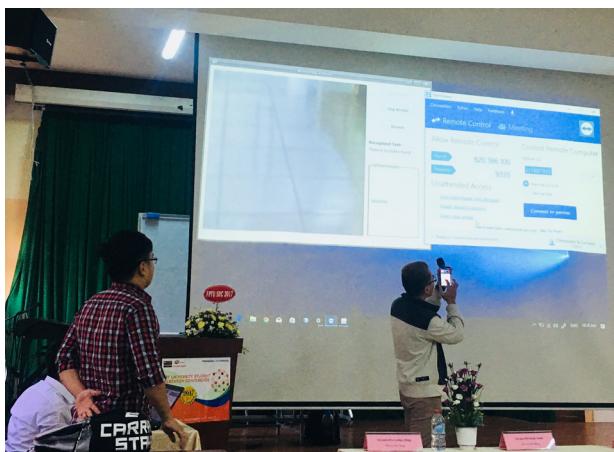
Team 5 was presenting subject: “FootBall Field Reservation System”



Team 3 was presenting subject: “Hacking and Security Techniques on Web Applications”



Team 6 was presenting subject: “Warning System on Parameters of the Devices in the Family”



Team 4 was presenting subject: “Applying iBeacon technology to build Automation Tollbooth Station”



Questions of the Scientific Council



**The contribution of
Mr. Lam Huu Khanh Phuong**



Tran Ngoc Tuan, PhD. awarded certificates and congratulated to the winning teams



Tran Ngoc Tuan, PhD. awarded certificates and congratulated to the winning teams



**Somebody take commemorative photograph
with winning teams**



**Posters were displayed at FPTU Student
Research Conference - Fall Semester 2017**



**Posters were displayed at FPTU Student
Research Conference - Fall Semester 2017**

The Student Research Conference Proceedings Vol. I & Vol. 2 were printed on May & August 2017. Readers can find in Library Department of HCMC FPT University and Ha Noi FPT University, also find it by electronic version on Dspace at link: <http://ds.libol.fpt.edu.vn>



HA NOI

Education and Training Zone -
Hoa Lac Hi-Tech Park - Km29,
Thang Long Highway, Thach
That, Hanoi
Tel. (+84-24) 73005588
Email: daihocfpt@fpt.edu.vn

HO CHI MINH CITY

2nd Floor, Innovation Building,
Block 24, Quang Trung Software
Park - Tan Chanh Hiep Ward,
District 12, HCMC
Tel. (+84-28) 73005588
Email: daihoc.hcm@fpt.edu.vn

DA NANG

137 Nguyen Thi Thap
Lien Chieu District, Da Nang
Tel. (+84-236) 7300999
Email: daihocfpt@fpt.edu.vn

CAN THO

160, 30/4 Street, An Phu Ward,
Ninh Kieu District, Can Tho City
Tel. (+84-292) 7303636
Hotline: (+84) 909070357
Email: fptu.cantho@fe.edu.vn