Chapter 5
A Survey on Detection and Analysis of Cyber Security Threats Through Monitoring Tools

Manjunath Kotari

Alva's Institute of Engineering and Technology, Moodbidri, India

Niranjan N. Chiplunkar

NMAM Institute of Technology, Nitte, India

ABSTRACT

Cyber crime is a serious threat for day-to-day transactions of the digital life. Overexposure of the personal details in social networks will lead to the cyber crime case. Therefore, detection and monitoring of cyber crime are challenging tasks. The cyber criminals are continually flooding the various intrusions all over the network. The cyber safety team should have a noteworthy challenge of filtering various such information. Continuous nonstop cyberattacks or intrusion examinations by security tools will significantly improve the threat alerts. However, cyber security becomes more expensive in the case of the above methods. The chapter provides systematic survey of various cyber security threats, evolution of intrusion detection systems, various monitoring mechanisms, open source cyber security monitoring tools, and various assessment techniques. The chapter also proposes a model of Cyber security detection and monitoring system and its challenges.

1. INTRODUCTION

Cyber security threats are major hurdle for the development activities of the Information Technology (IT) industry. The IT industry is facing severe crisis of cyber-crime activities in their business. A large set of data and assets of organizations are placed in cloud-based platform. The virtual cloud computing is facing various threats which include, Intrusions, Malwares, and Mining of Crypto currency. The

H.O.D.
Dept. Of Computer Science & Engineering
Alva's Institute of Engg. & Technology
Mijar, MOODBIDRI - 574 225

REVIEW OF INTELLIGENT SAFETY SYSTEM FOR WOMEN SECURITY

Sangeetha N A[1], Shetty Sonali[2], Shaima Abdul[3], Shruthi Kamath[4], Harish Kunder[5]

Department of Computer Science and Engineering

[1-4]Eight semester UG students, AIET, Moodbidri, Karnataka, India.

[5]Sr.Assistant Professor, AIET,Moodbidri,Karnataka,India.

nasangeethagowda@gmail.com,shettysonali1998@gmail.com,shaima1308983333@gmail.com,shruthikamath1998@gmail.com

ABSTRACT

In today's scenario, women harassment has become one of the major problems for women. A device has been proposed in which multiple devices and hardware consisting of a band like thing which can communicate well using a phone which is in turn connected to the web are integrated. The proper description of the design and working of a "Smart band" is been dealt in this paper. The components include trigger, micro controller (ATmega2560), GSM module (SIM900), GPS module (Neo-6M), IoT module (ESP-12E), NeuroStimulator, Buzzer and Vibrating Sensor. When a woman feels that she will be facing some trouble or she is already in such a situation, all she has to do is to hold ON the trigger present in the device. As soon as the device is activated, it will trace the current location using GPS (Global Positioning System) and also it will send the required message using GSM (Global System for Mobile communication) to the registered phone numbers and to the nearby police station. Here also the IoT module is used in order to track the location simultaneously so that it will be updated into the webpage. Neuro Stimulator helps to produce non-lethal electric shocks to detect the attacker, while the buzzer is used as an alarm to alert the nearby people.

Key words

GPS Tracker and GSM Module, IoT module, NeuroStimulator

1. INTRODUCTION

In Today's situation, the safety of women and girls is in tremendous danger specifically in India. The rate of crimes and the abusement happening against women is never less. But in fact it is increasing at a high rate especially harassing them every now and then, molestation, eve-teasing, rapes, kidnapping, domestic violence to the married women because of dowry. Many necessary measures have been adopted by the government to prohibit these misbehaving deeds but still it has not altered the growing rate of these crimes. The issue of sexual harassment in working areas is dynamically increasing. It is an unwanted behavior of some individual that causes discomfort to the other. Many such cases happen to women by men who work at high positions in a company or an organization. Women are also getting kidnapped at every 44 minutes, raped at every 47 minutes, 17 dowry deaths every day. Children face problems like child trafficking and kidnapping, while returning back to home from school. Having security applications for women, and a smartphone can help one send emergency information to selected people and letting them know if anything goes wrong. Sometimes there can be a situation where women had an accident in the late night while returning from work and there is no one around to help them. In such situations the person will not be able to deal with the situation if she is being attacked. And they might not be knowing the basic first-aid details. Nowadays indeed there are so many apps and devices evolved for women safety which can be made use only by a touch or one click or just by a shake.

2. EXISTING SYSTEM

The existing work proposes that it will inform the parents and the police about the current location of the victim. A GPS system is used to trace the present location of the victim and a GSM is used to send the messages to the predefined numbers. The work of pulse sensors is to check the pulse of the victim and in abnormal health situations like increase of the pulse rates the device also sends current GPS location to ambulance at every 10 sec in message format. Also a skeleton of a user friendly mobile application has been designed and presented named Women Empowerment which contains totally different laws associated with VAW and in addition to that it contains different health tips for women, which can help the rural as well as urban women to face any critical situation. There is also a new model for women safety, Once the switch is pressed the current location of women is collected and sent through GMS to the contact numbers registered in the Arduino, L293D is used to drive dc motor, in addition buzzer and shock system also used in this device. The system makes use of Arduino Mega 2560 as its main microcontroller. The system additionally generates a screaming alarm that uses real time clock, to generate an electrical shock to harm the assaulter for self-defense.

The work was so proposed that the user is accessible to quick and immediate help in any emergency situation. It makes use of GPS technology. The System refers to the Global Positioning System to trace out the position of the person and uses the Messaging service to send the message to get any kind of aid. When the switch is pressed the device will get activated. The audio and video recorder can now start to capture the live incident. If the pressure sensing element present in the device senses the physical pressure the messages are sent to contacts with the live recording through GSM module. Buzzer is provided within the device, so when the device gets triggered the buzzer produces high frequency sound in the surrounding. So that the people around there may hear the sound and come to the rescue of the victim. Spy-camera starts functioning when the device activates and captures the video of present location

H. O. D.
Dept. Of Computer Science & Engineering

Review on Android Application Security

P. Vignesh Pejathaya[1], Imran Khan[2], Vignesh Shetty[3], Sayeesh[4], Manish B. Shriyan[5]

[1]*Dept. of CSE, Alva's Institute of Engineering & Technology, Mijar, Moodbidri, India*
[2]*Dept. of CSE, Alva's Institute of Engineering & Technology, Mijar, Moodbidri, India*
[3]*Dept. of CSE, Alva's Institute of Engineering & Technology, Mijar, Moodbidri, India*
[4]*Dept. of CSE, Alva's Institute of Engineering & Technology, Mijar, Moodbidri, India*
[5]*Dept. of CSE, Alva's Institute of Engineering & Technology, Mijar, Moodbidri, India*

--***--

Abstract - *Smartphone's are used by billions of people all over the world which indicates the applications of the Smartphone is increasing and the main concern of any applications is the marketplaces to completely validate if an application is malicious or legitimate. Recent advances in hardware and telecommunications have enabled the development of low cost mobile devices equipped with a variety of sensors. It makes an easy target for the malware developers and other computer criminals. The anti-malware organizations and academic researchers have produced and proposed many security methods and mechanisms in order to recognize and classify the security threat of the Android operating system. This paper tells about the misuse of app permissions using Shared User ID, how two factor authentications fail due to inappropriate and improper usage of app permissions using spyware, data theft in Android applications, security breaches or attacks in Android and analysis of Android, and Windows operating system regarding its security.*

1. WORKING OF MOBILE APPLICATIONS

Mobile applications are at the epicenter of current development trends. Most of these applications have a client–server architecture. The client runs on the operating system, which is most frequently Android. This client is downloaded to the device from the app distribution platforms, where developers publish their wares. As perceived from the user's point of view, the client installed on the Smartphone is the mobile application. This is what the user interacts with to make purchases, pay bills, or read emails. But in fact, there is also another component: the server, which is hosted by the developer. Often this role is performed by the same software that is responsible for generating and processing content on the site. In other words, most often the server-side component is a web application that interacts with the mobile client over the Internet by means of a special application programming interface (API). Which gives us two types of vulnerabilities

1.1 Client-side vulnerability :

Android applications tend to contain critical vulnerabilities slightly more often than those written for iOS (43% vs. 38%). But this difference is not significant, and the overall security level of mobile application clients for Android

and iOS is roughly the same. About a third of all vulnerabilities on the client side for both platforms are high-risk ones

Fig -1

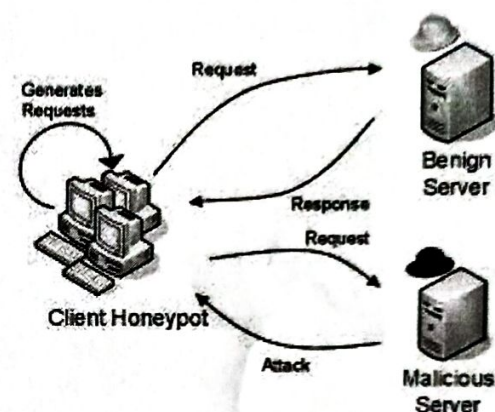### 1.2 Server-side vulnerability :

Server-side components contain vulnerabilities both in application code and in the app protection mechanisms. The latter include flaws in the implementation of two-factor authentication. Let us consider one vulnerability our experts encountered in an application. If two identical requests are sent to the server one right after the other, with a minimal interval between them, one-time passwords are sent to the user's device both as push notifications and via SMS to the linked phone number. The attacker can intercept SMS messages and impersonate the legitimate user, for instance, by cleaning out the user's bank account.

Fig -2

REVIEW OF AI CHATBOT

Prashanth S¹, Rakshith Gowda N², Sourabh Kakade³, Gouramma⁴, Vivek Sharma⁵

1,2,3,4Student, Alvas Institute of Engineering and Technology, Mijar Moodabidri.
5Assistant Professor, Alvas Institute of Engineering and Technology, Mijar Moodabidri.
---***---

Abstract - *A dialog-exchanging system generating a meaningful and empathetic conversation between human and a computer is called a chatbot which processes the natural language input that can be either a speech or a textual form and gives the response in the same language and expression as of the human. Various Natural Language Processing (NLP) techniques using the python library, Natural Language Tool Kit (NLTK) these inputs of speech from users can be processed and intelligent responses from such system engine can be gained to build up a human-like interaction. The chatbot is built with a limited dictionary but uses a great algorithm to imitate a real person. The chatbot can be used to find you an amusing partner and help you in bad times. If the users, ask the question to the chatbot it will give suggestion based on that question. The answers are appropriate to the user's queries, if the user find his answer to be invalid, it will give default message and will notify to admin. The algorithm used here has a intelligently built in logic and is designed to better chat with users.*

Keywords : Chatbot, Institution, Interactive method, Future studies, Educational domain.

1. INRTODUCTION

Conversational agent or Chatbot is a program that generates response based on given input to emulate human conversations in text mode. These applications are designed to simulate human-human interactions. Chatbots are predominantly used in business and corporate organizations including government, non-profit and private ones. Their functioning can range from customer service, product suggestion, product inquiry to personal assistant. Many of these chat agents are built using rule based techniques, retrieval techniques or simple machine learning algorithms.

In retrieval based techniques, chat agents scan for keywords within the input phrase and retrieves relevant answers based on the query string. They rely on keyword similarity and retrieved text is pulled from internal or external data sources including world wide web or organizational database. Some other advanced chatbots are developed with natural language processing(NLP) techniques and machine learning algorithms. Also, there are many commercial chat engines available, which help build chatbots based on client data input.

All Chatbots usually search for keywords, phrases, and examples that have been customized into their databases, yet some utilize more propelled strategies. In this paper the need for Chatbot in education domain and designed to provide user satisfaction.

2. LITERATURE SURVEY

Aditya Deshpande, Alisha Shahane, Darshana Gadre, Mrunmayi Deshpande et.al paper presents on survey of various chatbot implementation techniques. The research is based on the survey of various chatbot and it is shown that how the various chatbot differs from each other's. To make chatbot the various technology has been used to make. A chatbot can be considered as a question-answer system where experts provide knowledge for solicitation of users. A chatbot is a software designed to simulate a conversation with human partner. This survey paper aims to present an overview of an existing approaches of implementing a chatbot system. In this paper it as compares various chatbot from the first chatbot ELIZA to one of the latest chatbot like ALEXA, not only this is as shown various chatbots like IBM Watson, Siri, Tay, etc. It tells us about how it is implemented and how they actually work.

The design and implementation of several chatbots are developed a detailed survey of those systems. Mukesh Kumar, Sayali Hulawale, Sahil Pandita et.al paper presents on college management Chatbot. The college inquiry chat-bots will be built using artificial algorithms that analyse user's queries and understand user's message. The User can ask the question any college-related activities through the chat-bot without physically available to the college for inquiry. The System analysis the question and then answer it to the user. The user can ask any question related to college he asks about annual day, college fees, faculty details, sports day, etc. It helps the student to be updated what is happening in the college. In this the Chatbot usually remember the previous command in order to provide the functionality. So, whatever the user ask it will remember for it, it will help the Chatbot and it can analysis it for future references. If the Chatbot couldn't

Image Compression on Heterogeneous Images Using Convolutional Neural Networks

[1]Anix Jugal D'Cunha, [2]Sushant Mangasuli, [3]Vaibhavi Sahukar, [4]Afrah Saleem [5]Akshatha Sagar

Alva's Institute of Engineering and Technology,
Visvesvaraya Technological University, Mangalore, India

Abstract: Data compression has always been a key concern during transmission. Interests on image processing has been increased enormously from last decades. As a result different compression techniques has been introduced and purposed. Lossy image compression algorithms are used widely where some information gets lost during compression resulting high compression. However, we pay for their high compression rate with visual artifacts degrading the user experience. Deep convolutional neural networks (CNN) have become a widely used tool to address computer vision tasks very successfully.

I. INTRODUCTION

This chapter gives a brief introduction about the Machine Learning and project entitled "Image Compression Using Convolutional Neural Networks". With the development of multimedia technology and communication technology, multimedia entertainment has played an important role in people's daily lives. Pictures and videos take up the main part of multimedia entertainment. It brings austere challenge to store and transmit those data, and puts forward higher requirement on the limited-bandwidth internet, especially for large and high-quality digital images. The limited bandwidth of internet greatly restricts the development of image communication, and thus the image compression technology has been more and more aroused people's attention. The purpose of image compression is to represent and transmit the original large image with minimal bytes, and to restore the image with not-so-bad quality. Image compression reduced the burden of image storage and transmission on the network, and achieved rapid real-time processing online. The information of an image is fixed, but the different representations of the image lead to different changes in the amount of data stored in the image. So in the representation with larger amount of data, some data is useless or represent the information that is represented by other data, they are irrelevant or redundant. The main purpose of image compression is to compress the image by removing redundant or irrelevant information, and to store and transmit digital compressed data on a low bandwidth network. Image compression methods can be classified into two kinds: one may lose information during compression, and the other one can keep full information, that is, lossless coding methods and limited-distortion coding methods. Lossless coding methods will not suffer loss of information after compressing images, yet without a good compression ratio. The basic principle of this kind of methods is: an image consists of features, using the statistical features of the image, if a feature appears many times in the image, it will be encoded in shorter bits, and if a feature appears only once or limited times, it will be encoded in longer bits. And a complete image is always composed of a large number of repeated features. According to that, the image will be represented by many short-bits coding features and little long-bits coding features. On the basis of guaranteeing the image quality after compression, limited-distortion coding methods maximize the compression ratio. The original image and the compressed image looks very similar though some information has changed. The normal used limited-distortion coding methods are: the predictive coding method, the transform coding method and the statistical coding method. To replace the transform and inverse transform in traditional codecs, we design a symmetric CAE structure with multiple down sampling and up sampling units to generate feature maps with low dimensions. We optimize this CAE using an approximated rate-distortion loss function. The features of images can be learned automatically using deep learning models, rather than proposed manually. Suitable features can improve the performance of image recognition. Over the past years, features of images were always specified manually that depended on the designers' prior knowledge, and the number of features were very limited. Deep learning models can automatically learn unlimited number of features automatically. A good feature-extraction method is a prerequisite for optimization of image processing. In this study, we proposed a model to compress images. Based on AE, a multi-layer model is constructed. An image is put into the first layer and the output data from different level of layers reconstruct the original image in different level of comprehension. If the size of the output data from an arbitrary layer is less than the size of the original image, the representation in this layer is a compression representation. Because the model has more than one hidden layer whose neurons are less than the input layer's, the model can achieve multiple levels of features, and each level of features represent a compressed image. So, multiple compression ratio can be obtained using this model.

II. PROPOSED CONVOLUTIONAL AUTOENCODER BASED IMAGE COMPRESSION

The features of deep learning can be learned automatically using deep learning models rather than proposed manually. Suitable features can improve the performance of image recognition. Over the past years, features of images were always specified manually that depended on the designers' prior knowledge, and the number of features were very limited. Deep learning models can automatically learn unlimited number of features automatically. A good feature-extraction method is a prerequisite for optimization of image processing. In this study, we proposed a model to compress images. Based on AE, a multi-layer model is constructed. An image is put into the first layer and the output data from different level of layers reconstruct the original image in different level of comprehension. If the size of the output data from an arbitrary layer is less than the size of the original image, the representation in this layer is a compression representation. Because the model has more than one hidden layer whose neurons are

An Intelligent, Automated Toll Payment System

Aruna Kumari V [1], Brindashree B V [2], Chaitanya Tejas [3], Mithun Kumar D [4], prof.Hemanth Kumar
N P [5], prof.Arpitha J C [6]

Department of Computer Science and Engineering, AIET, Moodabidri
[1] arukushi06@gmail.com
[2] brindashreebv@gmail.com
[3] chaitanyatejas333@gmail.com
[4] mithunkumar745@gmail.com
[5] hemanthnp@gmail.com
[6] jcarpitha@gmail.com

Abstract— Automated Toll Collection System (ATCS) is one of the technologies that aims to achieve the Intelligent Transportation System's (TTS) which provides an efficient road and transportation infrastructure at the highways. The objective of this paper is to demonstrate a possible solution to overcome the problems encountered in manual toll collection systems like time wastage at the toll plaza, cashless transactions, over collection of toll charges, corruption on toll plaza, etc. This paper is aimed to provide an accurate and efficient ATCS based on a vehicle type classification method rather than the current implementation of toll collection that rely on sensor-based and human observation. To fulfil the aim, AI based desktop application will be used for extraction of vehicle license number which is the main input for the auto payment. This application consists of several sub components/modules such as image acquisition module, Image pre-processing module, vehicle extraction module, vehicle classification number plate identification and vehicle number extraction. The mobile app, which will be used by the truck drivers consists of a GPS interface module for getting the correct coordinates of the vehicle, in built database consisting of the coordinates of all the toll booths in the country, network/data transfer modules for sending the coordinates of vehicle along with other information such as vehicle number, vehicle type, etc. to a remote server. The web application facilitate auto payment based on the inputs from the AI application and the mobile app, enable the user to register their vehicle for auto payment with a payment gateway, view the transaction history, request for refund for wrong transactions, check the status of the refund, enable the toll booth user to generate bill for unregistered vehicle passing through the toll gate , create various reports, etc.

I. INTRODUCTION

With the recent advancements in technologies like Internet of Things (IoT), Artificial Intelligence (AI),etc, a lot of innovations have started coming in a number of application areas. One such application is an intelligent and automatic toll payment system for vehicle entry in toll plazas. This system makes lives simple both for the driver and the toll staff and clears unnecessary piling up of vehicles near the toll especially big trucks which create traffic jam. Even though it appears to be simple with many conventional solutions for this problem, many of them have their own limitations and warrants for trickier solutions. The existing National Electronic Toll Collection program of NHAI uses an RFID tag called FASTag that employs Radio Frequency Identification (RFID) technology for making toll payments directly from the prepaid account linked to it. It is affixed on the windscreen of the vehicle and enables to drive through toll plazas. Many banks such as ICICI, SBI, AXIS bank, etc. have already tied up with this program. FASTag is presently operational at 407 toll plazas across national and state highways. Even though this is a simple to use and efficient solution, this facility is provided for Car,

Jeep and Light Motor Vehicle (LMV) class only. More over the vehicles should pass only through the dedicated lanes equipped with the RFID reader for this to work. One of the major concerns now would be to allow vehicles especially big trucks to pass through any of the lanes, so that much of the traffic congestions near the toll gate can be minimized. Techniques available which could provide better security at the cost of very little time, also reducing the complexity of the design to a great extent. One of the issues with the current implementation of the toll collection system is human error may occur during the observation. Furthermore, as for the sensor-based and electronic tools based implementation, they are high in maintenance and need some configuration tools as in some of the existing works . Therefore, we proposed to overcome the issues by using ATCS that is camera-based (computer vision and combined with the advantages of machine learning). The advantages of this approach is low in maintenance and easy to be configured since only camera and a system are needed to classify the vehicle.

II. LITERATURE REVIEW

An RFID reader is a device that is used to commune with the RFID tag. The reader has a receiver that exhales radio waves. The tag should answer back by sending the data. GSM module is used for relating the entire system through the Internet. When the vehicle passes through the tollgate, the LCD system displays the information about that vehicle. The Motor Drive organizes the gate of the system. IR sensor is used to sense the presence of the vehicle. The alarm indicator shows the illegitimate crossing of vehicles through the tollgate. The paper proposes a system which includes Transponder, Antenna, Traffic Controller System and Central Server. The Transponder is the RFID tag which has a unique identification number. The Traffic Controller System builds the system differently from others. This system manages the allocation of incoming vehicles across a set of parallel toll gates. It makes use of Lane Allocation Algorithm. This algorithm allocates the gate entry for each of the cars approaching among the available set of toll gates. The system also utilizes traffic speed controller. Once the RFID tag reaches the range of the Toll Plaza, the Electronic Computer Unit is guided to reduce the speed of the vehicle.

When the vehicle crosses the sensors, which are fixed at a few meters distance from the toll plaza, tag will read through RFID Reader. The RFID tag has a unique identification number. The data read by the RFID reader will be sent to the micro controller. Then the micro controller processes the data for validation. If the consumers are valid, then data is sent to the processing unit.

1

Performance Comparison of a Parallel Recommender Algorithm across three Hadoop-based Frameworks

Aishwarya J[1], K Manasa[2], Manikya K[3], Pallavi sutar[4], prof. Hemanth Kumar N P[5]

Department of Computer Science and Engineering, AIET, Moodabidri

[1]*aish.nv15@gmail.com*
[2]*mk3954479@gmail.com*
[3]*manikyak1998@gmail.com*
[4]*pallavisutar1998@gmail.com*
[5]*hemanthnp@gmail.com*

Abstract—One of the challenges our society faces is the ever increasing amount of data. Among existing platforms that address the system requirements, Hadoop is a framework widely used to store and analyze "big data". On the human side, one of the aids to finding the things people really want is recommendation systems. This paper evaluates highly scalable parallel algorithms for recommendation systems with application to very large data sets. A particular goal is to evaluate an open source Java message passing library for parallel computing called MPJ Express, which has been integrated with Hadoop. As a demonstration we use MPJ Express to implement collaborative filtering on various data sets using the algorithm ALSWR (Alternating-Least-Squares with Weighted- λ-Regularization). We benchmark the performance and demonstrate parallel speedup on Movie lens and Yahoo Music data sets, comparing our results with two other frameworks: Mahout and Spark. Our results indicate that MPJ Express implementation of ALSWR has very competitive performance and scalability in comparison with the two other frameworks.

Index Terms—HPC, MPJ Express, Hadoop, Map Reduce, YARN, Spark, Mahout

I. INTRODUCTION

Over the last decade Apache Hadoop has established itself as a pillar in the ecosystem of software frameworks for "big data" processing. As an open source, mostly Java-based Apache project with many industrial contributors, it retains a commanding position in its field.
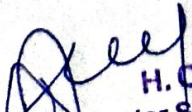
When first released Hadoop was a platform primarily supporting the Map Reduce programming model, and other projects built on top of Map Reduce. Around 2014 with the release of Hadoop 2.0 the platform was re-factored into a separate YARN (Yet another Resource Negotiator) resource allocation manager, with Map Reduce now just one of multiple possible distributed computation frame- works that could be supported on top of YARN. Several other major big data projects rapidly mi- grated to allow execution on the Hadoop YARN platform (for example APR pache Spark [24], Apache Graph [1], Apache Tez [15], and Microsoft Dryad [9]). Around the same time the present authors envisaged adding our existing MPJ Express frame- work for MPI-like

computation in Java to that distinguished group, and developed a version of our software that could also run under Hadoop YARN [22].

MPJ Express is a relatively conservative port of the standard MPI 1.2 parallel programming interface to Java, and is provided with both "pure Java" implementations (based on Java sockets and threads) and "native" implementations exploiting specific interconnect interfaces, or implementations on top of standard MPI. The vision was thus to support MPJ as one computational framework among many largely Java-based or JVM-based frameworks that could be mixed and matched for different stages of complex big data processing, with Hadoop and HDFS (the Hadoop Distributed File System) as the "glue" between stages.

The main goal of the present paper is to pro- vide evidence that such a scenario can be realized and that it may be advantageous. We concentrate on one particular computationally intensive "big data" problem - generating product recommendation through the collaborative filtering algorithm ALSWR (Alternating Least Squares with Lambda Regularization). A version of this algorithm was developed and evaluated using MPJ running under Hadoop. We then go on to compare our implementation with two existing implementations of AL- SWR that can run under Hadoop— one taken from the Apache Mahout project using Map Reduce, and one using Apache Spark. Results suggest the MPJ approach can provide useful performance gains over these other established Big Data frameworks on suitable compute-intensive kernels.

The rest of the paper is organized as follows. Section I-A reviews selected related work. Back- ground materials in Section II review Hadoop, YARN and HDFS; outline the architecture of MPJ Express and its integration in YARN; and give an overview of the collaborative filtering technique. Section III describes how we implement the collaborative filtering with ALSWR in MPJ. The Section IV evaluates and compares our results with Mahout and Spark. Section V concludes the paper and discusses future works.