Portfolio Summary

Balaji Chandrasekaran
School of Computing, Informatics and Decision Systems
Arizona State University
Tempe, Arizona
bchandr6@asu.edu

During my tenure at Arizona State University for pursuing my dream as a master's student in Computer Science, I did several outstanding projects and acquired high level of knowledge with par to the ever-changing computing paradigm. It was a tough decision to decide three of the several projects that I did. I choose the following projects for my portfolio – 1) "GeoSpatial Data and Spatio Temporal Analysis" which was done in the course CSE 512 - "Distributed Database Systems" under professor 'Mohamed Sarwat' during Fall'16, 2) "Client Server Message Passing" which was done in the course CSE 531 - "Distributed/Multiprocess Operating System" under professor 'Partha Dasgupta' during Fall'15 and 3) "Social Computing and Information Assurance" which was done in the course CSE 543 - "Information Assurance & Security" under professor 'Stephen S. Yau' during Spring'17. Here is a summary about every project that I am attaching in my portfolio below. The detailed information regarding the project is provided in the individual report of each project.

- GeoSpatial Data and Spatio Temporal Analysis

  Big data is a field of ever growing interest for researchers. Big data is so important that when used properly it could pave way to several medicinal innovations. This project deals with big data and provides us with knowledge and insight into various technologies that are used in it. This project provides a hands-on experience into several technologies like Hadoop, Spark etc which are of prime importance when handling large distributed data. The project was to find the top 50 hotspots in the envelope which was in New York with given details regarding taxi data at that location. In this project, we had to used Spatio temporal analysis which is another concept used in industry for several purposes when it comes to big data. The project was very close to industrial standards.

- Distributed/Multiprocess Operating System

  The project was related Operating System. The main goal of the project was to develop a message passing system which would help multiple threads to communicate with each other in a shared memory space. The project was subdivided in several phases so that we can understand in detail about the various important building blocks of a message passing system. The first phase of the project was to develop a scheduler since the project involves multiple threads, there should be a scheduler present so that the multiple threads can be scheduled. For this purpose, we used the ucontext library which is already available in C and developed the scheduler for scheduling purposes. The second phase of the project involved the process of creating a way to handle multiple threads without the problem of critical section. To solve this, we developed a synchronization mechanism in which we had to define and develop semaphores and semaphore related functions for synchronization. The third phase laid the foundation for developing a message passing system in which we create thread control blocks and define a message passing system using the previously created libraries in the phase 1 and 2 for the various purpose of scheduling and synchronization. The final phase of the project was to combine all the previous phase and build a client server application which worked using the message passing system.

- Social Computing and Information Assurance

  This was a research project whose purpose is to learn about the significance of information assurance and the consequences of information theft. This project provided us with an opportunity to go through several previously written papers on security and several case studies involving security theft. This opened our knowledge paradigm wide open. It was fascinating to see the importance of information and consequences of stealing the same via several case studies. The project thought us on how to protect information and the role of ourselves as a common user. The topic that we dealt with was social computing which is a growing and important paradigm to study about since it is part of the life that we live in. The conclusions and recommendations made us into a better developer as it gave us great insight on how to prevent several threats while developing products.

GeoSpatial Data and Spatio Temporal Analysis

Balaji Chandrasekaran
School of Computing, Informatics and Decision Systems
Arizona State University
Tempe, Arizona
bchandr6@asu.edu

***Abstract*—The project comprises of several phases. The initial two phases constituted of familiarizing GeoSpark, the implementation of GeoSpark for GeoSpatial Objects and running a statistical analysis of the distributed techniques such as Spatial Range Query, Spatial KNN Query, Spatial Join Query with and without the join of PointRDD using R-tree or Grid. The analysis was done using Ganglia and the system constituted of Hadoop Cluster that was running spark shell with Geo-Spark jar. The final phase of the project constituted of Spatio-Temporal analysis of a subset of NYC Yellow taxi cab dataset to compute the top 50 hotspots in the given envelope using Getis-Ord statistic.**

Keywords—Hadoop; Spark; GeoSpark; Getis-Ord; R-tree; Ganglia; Spatio-Temporal Analysis;

I. INTRODUCTION

Big Data comprises of extremely large data sets that could be used for analyzing, which results in identifying several meaningful patterns and trends. This process of studying huge amount of legacy data is of prime most significance for predicting the current and future trends. Most of the large data sets are constituted into clusters which helps them in maintaining them and processing the same. Even though there are plethora of tools available for monitoring and maintaining the clusters almost all organizations fail to hit the balance which would result in efficient and effective way of processing big data clusters and obtain desired result. This project deals with providing a hands-on experience of the Hadoop and Spark clusters, understanding the concepts of MapReduce, performance variations due to R-tree index and grid. The project also provides us with a real-world application of Apache Spark and Hadoop clusters for distributed computing with the help of GISCUP 2016 problem solving. [2]

II. PROJECT DESCRIPTION

A. Phase I: System Implementation Plan/System Prototype.

This phase involved the process of setting up of the Hadoop clusters and understand the System behavior by doing several functionalities.

1. Setup of Hadoop Clusters and load the GeoSpark jar into Apache Spark Scala shell.
2. Operations to be done:
    a. Creation of the PointRDD.
    b. Spatial Range Query: Query the PointRDD with and without building R-tree index using given input.
    c. Spatial KNN Query: Query the PointRDD with and without the R-tree index using given input.
    d. Spatial Join Query: Creation of rectangleRDD and use the same for joining pointRDD with the help of equal grid and with/without R-tree Index, R-tree grid and R-tree Index.

B. Phase II: System Demonstration/Experimental Evaluation and Analysis of Results.

There were two halves to this phase. First half was to implement a JAVA function using the Cartesian product algorithm for the purpose of Spatial Join Query. For the same use the GeoSpark Spatial Range Query that is already present in GeoSpark.

The other half was to compare the various functionality that we did in phase I and provide a valid reasoning for the differences that was observed between them using Ganglia which is a cluster monitoring tool and based upon the statistics of Execution time, Average memory, Average CPU utilization of the cluster.

C. Phase III:GISCUP 2016. [2]

In this phase, we took the problem that was provided in GISCUP 2016 competition which was a real-world application of distributed computing. The main objective of the problem was to find the top 50 hotspots of the given NYC taxi cab dataset using the Getis-Ord correlation.

This was the final phase and one of the prime most phase of the project. As we discussed earlier this phase gives us insight on how big data is used in real world and how it can be used to create a better living. The solution of this phase comprises of various application which are from business oriented to elevate the living of public by using it in Urban planning, business startup, transportation management etc. Hotspots are nothing but locations in the given envelope (boundary under consideration) which constitutes of greater traffic/activity. The data involved in

this phase is both temporal (periodic) and as well as spatial which means there would have to be a massive dataset for computation. We use the MapReduce capabilities of Hadoop and Spark Distributed-ness to resolve the problem.

III. Result Description

A. *Phase I: System Implementation Plan/System Prototype.*

The phase I was the initial phase whose main objective was to get to know the Hadoop environment. In this phase, we setup Hadoop and then Apache Spark on it and performed the operations mentioned in the problem statement in previous section. The setup constituted of 3 physical machines one being master and the other two being slaves running Ubuntu, Hadoop 2.6.4 and Spark 2.0.1. The result of the phase was a demo video which explains and goes through the process of the initial phase and the outcome was knowledge on how to run Hadoop clusters and then Apache spark on them and performing operations on spatial data in the environment that has been setup. [4]

TABLE I. Cluster Setup

Cluster Setup			
System	**Memory**	**CPU**	**Cores**
Master	2. 9 GB	Intel i5	2
Worker1	6.7 GB	Intel i7	4
Worker2	6.7 GB	Intel i5	2

B. *Phase II: System Demonstration/Experimental Evaluation and Analysis of Results.*

The phase II was an extension of the previous phase. The objective of this phase was divided into two, first objective being to understand in depth the several variations of spatial operation that were carried out on the environmental setup in previous phase. This also comprise of a way to justify the deviations in variations of same operation done in previous phase, for this objective we choose Ganglia as the monitoring tool for Hadoop cluster monitoring. We studied the variations using several metrics like memory and CPU usage, execution time and presented our conclusion in a detailed manner. One of the prime observation is that — Queries those use indexing is faster compared to the ones that does not use indexing, but they consume more memory and nested loop queries with grid partitioning are faster than Cartesian join queries. The other Objective of the phase was to implement the spatial join query for which we used simple Cartesian product algorithm present and GeoSpark API.

C. *Phase III:GISCUP 2016. [2]*

The phase III problem has been already discussed in an elaborate fashion. For us to implement a solution to the given problem we considered several ways of solving the problem. The data that was given with the problem set was too huge to be processed due to the limited infrastructure available, so we filtered it down to January 2015 – 31 days. Getis-Ord was used since it is the most standard and trusted one which is used for denoting patterns in spatial type data clusters with the help of z-score that it generates. The entire envelope that was given for consideration can be divided into smaller regions of unit length which is a region of pickup or drop operation. These smaller regions can be represented as cells. We then use these cells and identify the top 50 hotspots with the help of total number of pickup operations carried out inside each cell.

We use the Getis-Ord statistic to calculate the score for each of the cell. Which is then used for identifying the top 50 hotspots.

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{[n\sum_{j=1}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2]}{n-1}}},$$

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2}$$

G_i^* – Getis-Ord, $\bar{x}$ – mean of cell scores, S – standard deviation, n – no of cells, x_i – value of cell, $W_{i,j}$ – neighbor parameter of cell.

Fig. 1. Getis-Ord Equations.

After doing in depth research we choose to follow the approach that has been mentioned in the following paper "Spatio temporal hotspot computation on Apache Spark" by Paras Mehta, Christian Windolf, Anges Voisard. [3] The algorithm that we used consisted of four main steps –

1) Data Pre-Processing: In this we load the given data into HDFS by processing the CSV file and removing unwanted columns i.e. fields and removing the data which are outside the given envelope (40.5, –40.9, -74.25, -73.7) i.e. outliers.

2) Cell Generation: In this step, we created the cells and created a mapping that could be used for the Getis-Ord calculation for every cell based on a naïve approach using the neighborhood cells. Each of the cell created was typically 0.01 latitude high and 0.01 longitude wide.

3) Score Calculation: For score calculation, we do a MapReduce. Each cell that is under consideration for which Getis-Ord need to be calculated can be surrounded by either 26/17/11 neighboring cells depending upon whether it is inside/boundary/corner respectively. The Map phase consists of assigning a value of 1 and key which consists of coordinates of Spatio-temporal type to the cell that

localizes each point of activity in spatial and temporal planes. The Reduce phase comprises of mapping key with value in which value is consolidated cell score for all the cells. The various partitions are unaware of each other's data as this is done inside each of the RDD.

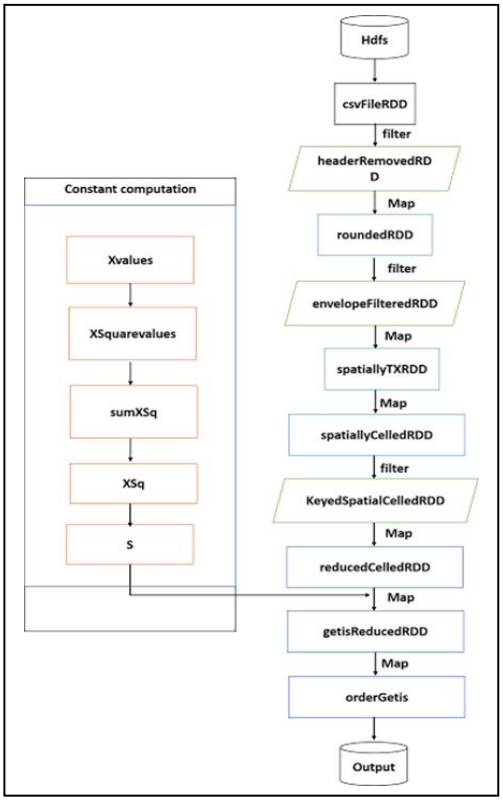

Fig. 2. Flow Chart.

4) We then used a HashMap and stored <KEY, VALUE>:<Cell-Value, Getis-Ord Score>. The HashMap was then sorted into a list with respect to the Getis-Ord Score, this is the shuffle phase of the solution. Then we retrieve and provide the top 50 hotspots from the list which are the required ones.

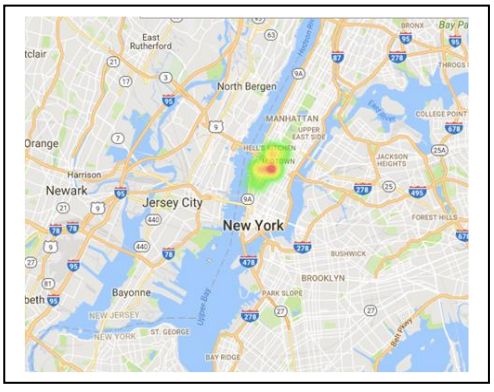

Fig. 3. Heat-Map visualization of top 50 Hotspots.

IV. Contribution

A. Cluster Setup

One of the machine in the cluster was setup and maintained by me. My Laptop enacted the role of Worker1. The process of setting up Hadoop and extending it to run Apache Spark was done by me. Apart from that I was also responsible for the process of integrating it with the master.

B. Video

Video management was done by me. I was responsible for recording and editing the video which was the outcome of phase I.

C. Development

I was responsible for the development of the spatial range query with/without indexing and the spatial KNN query with indexing. This provided me with great insight into spatial operations in Big Data. When it comes to phase III, the process of data preprocessing and cell generation were my responsibility.

D. Algorithm

The solution was constructed after careful consideration of several algorithms/techniques available. I made sure that I have full knowledge of the algorithm specified in "Spatio temporal hotspot computation on Apache Spark" [3] and about Getis-Ord statistic. [7]

E. Ganglia

When it comes to ganglia monitoring I was responsible for collecting data and necessary parameters about my machine (Worker1). This data was then used for constructing the report for phase II, which constituted of comparing the different types of operations that were carried out in previous phase.

F. Report

I was responsible for spatial mapping of hotspots, flowchart, tables, bar graphs, references section, abstract section and introduction section.

V. Knowledge Acquired

A. Skills

Hadoop, Apache Spark, MapReduce, Spatial Operations (Queries – join, range, KNN), Ganglia, HDFS, Java, Scala, GeoSpatial API, R-tree, RDD – Resilient Distributed Datasets, Getis-Ord statistic.

The project also provided us with a practical and real world application of Hadoop – Spark Clusters and MapReduce. The application provided us insight on how important and un-separable Big Data is in real life scenario. It was really fascinating to see such an example which provided us a base on how Hadoop could be used in several real-life scenarios for the betterment of life.

VI. TEAM MEMBERS

Aravind Rajendran, Purushotham Kaushik Swaminathan, Suresh Gururajan, Thamizh Arasan Rajendran.

REFERENCES

[1] Ord, J.K. and A. Getis. 1995. "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application" in Geographical Analysis 27(4).

[2] ACM SIGSPATIAL Cup 2016, Problem Definition (Online): http://sigspatial2016.sigspatial.org/giscup2016/problem Accessed: 03/24/2017

[3] Mehta, Paras, Christian Windolf, and Agnès Voisard. "Spatio-Temporal Hotspot Computation on Apache Spark (GIS Cup)."

[4] Group 25 - DDS Project Phase 1, submission link: https://www.youtube.com/watch?v=ShfpMlbMFYo

[5] Jia Yu, Jinxuan Wu, Mohamed Sarwat. "GeoSpark: A Cluster Computing Framework for Processing Large-Scale Spatial Data". (short paper) In Proceeding of the ACM International Conference on Advances in Geographic Information Systems ACM SIGSPATIAL GIS 2015, Seattle, WA, USA November 2015

[6] The Scala Programming Language, https://www.scala-lang.org/

[7] Getis Ord Statistic for Geospatial operations: https://en.wikipedia.org/wiki/Geospatial_analysis

[8] Data Serialization (Online), accessed 03/24/2017, https://spark.apache.org/docs/latest/tuning.html#data-serialization

[9] Zaharia, Matei, et al. "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing." Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. USENIX Association, 2012.

Client Server Message Passing

Balaji Chandrasekaran
School of Computing, Informatics and Decision Systems
Arizona State University
Tempe, Arizona
bchandr6@asu.edu

***Abstract*—The project is made up of four phases. Each phase is interconnected and current phase depends on previous phases. In every phase, we build a feature which will all be ultimately used in the final phase to build our message passing program. The project is developed to account for the parallel computational environment. The three main concepts that it deals with is Multi – Threading, Synchronization and message passing. For achieving all these three concepts we develop a Scheduler, Semaphore and message passing interface respectively which comprises of the first three phases of project and the fourth phase deals with development of a client server system which comprises all the features those developed previously.**

Keywords—Message Passing; Scheduler; ucontext; Thread Synchronization; Semaphore; Multi-Threading; Mutex;

I. INTRODUCTION

The main objective of the project is to learn about three main concepts which are threading, synchronization and message passing Interface which is used for parallel applications. In order to learn the above concepts, this project was done in total of four phases in which we did a thorough study of each of the concepts. We used C language and developed this project. In the first phase, we learnt about multi-threading for which we developed a library for generating user level multithreads using the ucontext library that is available. The second phase consisted of developing a mechanism that could handle synchronization of the threads for which we used semaphores. We then used the semaphores that is synchronization and solved the reader's – writer's problem in this phase. The third phase consisted of creating a message passing system which used the result of the previous phases. The fourth and final phase consisted of creating a client server machine which was built over the message passing system. The project gave us in-depth knowledge and understanding of threads working mechanism. It also thought us the importance of synchronization and how message passing takes place in a parallel architecture without affecting its purpose.

II. SOLUTION

A. Phase I : Scheduler for User level Multi Threads.

The first phase deals with the process of implementing a user level multi-threading. For the same purpose, we used ucontext library which is available. Uncontext stands for user context which consists of four important methods getcontext, setcontext, makecontext and swapcontext. We used three of these functions which are swapcontext, getcontext and makecontext. Getcontext function takes in a single parameter regarding the current context which is of struct type of ucontext_t which consists of four members which are "pointer to context that will be resumed – ucontext-t", "stack used by the context – stack-t"," blocked signals set - sigset-t", "machine specific representation of context – mcontext-t". The getcontext function sets its ucontext_t parameter with the current user context of the function/thread that calls the getcontext. Makecontext takes in a context, a function and several other arguments, it is used for modifying the context of the context present as a parameter of the function. When there is a call to the context again in the future the call is passed to the function that is present in as a parameter in the makecontext call. The swapcontext function takes in two parameters which are both context type and then swaps both of their context so that the running context is changed. In this program of phase I, we had three libraries that we had implemented and used for creating multiple threads. The first library was the "tcb" library which had the implementation of thread control block. Thread control block is nothing but a data structure which contains information regarding a thread to manage or control it properly. Some of the main information that are required to manage a thread would be its identifier, PC (program counter), Pointer to thread stack, thread state, pointer to PCB of its corresponding process. For obtaining and storing this information we used the ucontext library which consists of several functions as we have seen already for implementing TCB easily. The library also consists of a function init_TCB which is used for making context of threads and assigning the same to TCB type objects. Then we had the "q" library which had the implementation of creating and managing a circular doubly linked list that is a Queue. This Queue acted as a scheduler buffer which consisted of current and other threads that are to be executed in the order of execution. The library had InitQ, RotateQ, AddQ and DelQ function for initializing the queue, rotating the queue head, adding into the queue and deleting a queue node respectively. "threads" library is responsible for thread functionality like start_thread, run and yield function. Start thread makes call to init_TCB in the "tcb" library for initializing the context and adds the TCB to the queue. Run function gets the context and swaps it with the next context in the queue. Yield function makes a call to RotateQ for rotating the queue head and swaps the context to the next context in the queue. Then there is the testing program which has the main in it and consists of two functions namely add and mul. The purpose of the function is to do some mathematical

operation and these function act as the context to the two threads that we are going to create and make them run.

B. Phase II: Semaphores and Reader's – Writer's Problem.

The second phase consisted of the process of implementing synchronization into the multi threads paradigm. We then used the implementation of this synchronization and solved the Reader's – Writer's problem. Synchronization is something that is very important when it comes to multithreaded paradigm and if not handled properly it could lead to undesirable consequences. The main objective of synchronization is to solve the problem of critical section. The critical section problem arises due to the multi-threaded environment. When say there are two processes which are trying to access the same memory and these processes are run in parallel then there could be situation when both the process could access the memory at the same time and destroy the meaningfulness of the data. This kind of situation is known as critical section problem and the shared memory related code is known as the critical section of the code. We need some kind of synchronization that is understanding between the processes accessing the critical section that only one can access such section of code at a time and the other process as to wait for the current process to complete execution. For solving this problem there was a protocol developed which had three basic steps to solve the problem:

- Entry: A process entering critical section must make sure no other process is accessing the same critical section.
- Critical Section: After gaining access and entering the critical section it can execute the section and other process requesting the section must wait for it to complete.
- Exit: The process leaving the critical section specifies the same and clears its access so that another process can access the section.

The problem implementing such a protocol in an effective way is a challenging task. For this task, we implemented semaphores. Semaphore are nothing but variables which will be used for controlling the access of the critical section in parallel processing multi-threaded environment. In this phase, we used all the libraries from the previous phase so that we can use the multi-threaded scheduler in the current phase as well. We then created a library called as "sem" which contained the implementation of the semaphore, it had three functions which are "CreateSem – which was used for creating the semaphore, P – which is used for wait functionality of a semaphore", "V – this function is used for signal functionality of a semaphore". For achieving synchronization, we use the signal and wait function by checking the semaphore variable for granting permission that is signaling to use a critical section or making the thread to wait in the queue. Using this library, we solve the reader's – writer's problem which uses three semaphore variables namely reader semaphore, write semaphore and mutually exclusive semaphore. Mutually exclusive semaphore ensures the calculation for wait and signal of various threads gets executed without the problem of critical section that is there can be only one thread in entry or exit section at a given time. The reader semaphore corresponds to the reader functions which are trying to read a critical section and the writer semaphore variable account for all the writer functions which is trying to write in a critical section of memory. The main issue with a reader writer problem is that when there is a share of critical section given to a reader there should be no other reader that is coming in for critical section should be kept waiting. Using the above mentioned three semaphore variables and the function in the "sem" library we resolve the reader writer problem.

C. Phase III: Message Passing Interface.

The objective of this phase was to develop a message passing system which was multi-threaded and synchronized. For this phase, we utilized the result of the previous two phases. The objective of a message passing system is when we run a process on a parallel system we should have definitive way for communication between the process that we execute in that architecture and they should be able to communicate with one another by sharing information of each other. In a message passing system there is no way two process can share same memory, they look to do so but they don't actually do it but rather have their own copy of variables to work with. We developed a library called as msg using the previous libraries. The objective of the library is to implement message passing system. We defined the port structure which contains a message which is a two-dimensional square array of size 10, each port contains three of the semaphores namely sem_send, sem_recv and mutex like the reader's – writer's problem we have seen in the previous phase. The purpose of sem_send and sem_recv is to take care of the synchronization at the port like the writer semaphore and reader semaphore respectively and muter semaphore for controlling the entry and exit sections of synchronization mechanism that is the P and V function. We declared array of ports of length 100 and used the same for message passing. We implemented three functions which constituted the two main functions of message passing which is send and receive, the third function was init function which was used for initializing the ports. In the send function, there would be the wait function (P) from the sem library which makes the process to wait until there is a signal of critical section being free. The process of changing the semaphore variables that is the sem_recv and sem_send in the send function acts as a critical section which is being taken care of by the mutex semaphore. Similarly, there is the receive function which handles receiving of the message sent by the port using the semaphore variables. Later these ports are used in the next phase to develop a client server system for communicating using message passing Interface that we built.

D. Phase IV: Client – Server Architecture.

We used the previous phase output which is the message passing interface and developed a client server architecture like system which would be able to communicate with each other using the message passing system. This phase we used the msg library and created a program which consisted of two function which are client and server. The client sent messages to the server, the messages sent was in a random fashion using

a variable with modulo of three which was chosen randomly, three being the number of cases present in the client. The server receives the messages and stores them and sends back to the respective client from it. This way clients communicate with each other or with the server in this architecture. The client server functions act as the multi-threaded environment in this program and there must be certain synchronization mechanism which is being taken care of by the msg passing interface that we developed in the previous phase using the synchronization mechanism. These multi-level threads communicate with each other by passing messages. In this program, we created three client threads and one server thread which acted as the middle man for communication. It was very interesting to see how a message passing interface works at a very low level in the Operating System.

III. Results

A. Phase I : Scheduler for User level Multi Threads.

The result of the first phase was a library that would help us in implementing multiple threads for which we had to also implement a scheduler for handling the multiple threads. The scheduler was nothing but a queue. The result of this phase consisted of three libraries namely the "tcb", "q" and "threads". The tcb was for implementing the thread control block, q for the scheduler queue and threads for implementing the threads.

B. Phase II: Semaphores and Reader's – Writer's Problem.

The phase II result had an extra library to the already created libraries which provided the synchronization mechanism that is they provided an implementation of semaphores and their synchronization functions which are wait and signal function. The name of the library was sem. Using this library, we then created a program for resolving the reader's and writer's problem using the libraries.

C. Phase III: Message Passing Interface.

Result consisted of another library added to the previously created libraries which had the feature of message passing system. This library known as the msg library had the implementation of a message passing interface using the synchronization and scheduler that we already created. It had the two-main routine of send and receive for message passing between the processes or threads.

D. Phase IV: Client – Server Architecture.

The output of this phase had a program which consisted of the client server architecture using the message passing interface that we created in the previous phase. The program invoked three client threads and one of server thread. This paved way for the communication of the client and server once the threads where started. The result of this phase is that there is no guarantee which client's thread message gets delivered first. But when it comes to multi-threaded parallel computing paradigm the problem of resolving the critical section is solved by this system.

IV. Knowledge Acquired

The amount of knowledge that this project has thought us is enormous. It gave us profound knowledge on several aspects of multi-threaded environment and parallel computing paradigm.

A. Phase I : Scheduler for User level Multi Threads.

1) Multi Threads: This phase thought us on the understanding of multi threads. It gave us knowledge about the drawbacks or consequences of using multi threads.

2) Scheduler: This phase provided us with great insight on scheduler and how they work and how they actually schedule various jobs and maintain the queue. By implementing a scheduler on own this phase also provided us with practical experience of developing our own scheduler.

B. Phase II: Semaphores and Reader's – Writer's Problem.

1) Semaphores: This phase thought on resolving the drawback of using multi-threaded environment that we saw in previous phasse by usisng semaphores. It helped us to learn on how a semaphore can erradicate the problem of critical section and how to resolve any futher problem of implementing semaphore properly.

2) Reader's - Writer's : The practical example of solving a reader's – writer's problem helped us to a great extent on learning the implications of semaphores and how to properly implement one without the critical section problem.

3) Synchronization : This phase also thought us about the importance of synchronization and how it is achieved with the help of semaphores.

C. Phase III: Message Passing Interface.

1) Message Passing : In this phase we learnt about message passing and its importance in a parallel computing paradigm.

2) We also implemented a message passing system on our own that gave us practical knowledge on how to build a message passing system.

D. Phase IV: Client – Server Architecture.

1) This phase we built a client server architecture which gave us practical application of a message passsing system and how it works.

2) We saw how synchronization is used in realtime scenario.

The project also gave us lots of knowledge on implementing our own mechanisms for synchronization, multi-threading and message passing. It thought us about ucontext library in C and how it works. It provided us with in-depth understanding of C on pointer handling. It also gave us insight on how C is the perfect language when it comes to operating system related concept.

V. TEAM

Kandhan Sekar, Vimal Khanna Vadivelu

Social Computing and Information Assurance

Balaji Chandrasekaran
School of Computing, Informatics and Decision Systems
Arizona State University
Tempe, Arizona
bchandr6@asu.edu

***Abstract*—The interaction of people with computers are increasing exponentially day by day. It paves way for various kind of threats and cheating due to the increased interaction of humans with their gadgets. Social Computing is a paradigm that deals with interaction of users with each other through a computing device. Many a time the social computing is the target of hackers. In this paper, we deal with various sub divisions of social computing and study each one of them in depth. We also try to find various weakness present in it and propose solutions to them if there is any. We also talk about the preventive measures that could be taken for securing information and study several case studies to achieve the same.**

Keywords—Social Computing; Bluetooth; Wearables; Social networking; video games;

I. INTRODUCTION

Social Computing is the new paradigm of computing that is increasing exponentially day by day. Each user spends hours of their time in the social computing paradigm one way or another. The social computing paradigm comprises of many things including Online Gaming – Steam games, Warcraft, Dota 2, Social Networking – Facebook, Twitter, Instagram, Snapchat, Wearables – Smartwatch, glass, VR etc. There several other social computing sources as well, in this paper we are going to elaborately study these three sub categories. As we can see form the above example of social computing, it is something that is not going to perish and as long as human life exists on this planet there is going to be an exponential increase of users in this paradigm day by day. As the users increase the number of peoples that is the hackers who are trying to exploit the environment also increases. Information is the new wealth of humans and the new currency. Information can either make you a billionaire in overnight or bankrupt your business. As we humans always do when it comes to matter of such important things it is necessary in this case also to thorough study of the information theft and we should take all the necessary steps to protect our information as it is a matter of life or death. We will study in depth of the various mechanisms that could be used for stealing information and analyze each one of them. In wearables, we study about the security of the medium of communication, security of the device itself and role of individual in protecting his/her information. In social networking, we study regarding various spams, user pro activeness, user prevention measures, we talk about the various ways users can be cheated in Social gaming, the need for protecting the servers and identity theft. We study several case studies like the second life [1], Age of cannon [2], Polaris Bot [3], AFS Software Card Shuffling Algorithm [4], Absolute Poker Insider Attack [5]. We acquire certain knowledge from these case studies and use them to prevent future mishaps.

II. SOLUTION EXPLANATION

The introduction we dealt with three paradigms that we are going to do in depth study about. All the three paradigms are affecting the lives of humans one way or another. The prime most important thing is that all these paradigms are growing day by day and they are made up of information that is always at the risk of being attacked by hackers. We should deal with the various threats that the paradigms face and the concerned creators should also be proactive in eliminating future threats.

We will do a thorough security assessment on each of the three environments that we will be studying and provide with some feedback on the same. We will then be providing some of counter measures to mitigate those threats and give us a view of what are the root causes of them. We will also be providing to users with some information about the pivotal importance of their role in these environments to safeguard their information.

A. Online Gaming Industry.

Online gaming industry is something that makes up the life of several individuals. I myself am aware of many individuals whose life is all about gaming. There is a multibillion dollar industry revolving around the online games platform. There are several different famous online multiplayer games whose main objective is to provide a social environment and culture which makes people to enjoy themselves. We all would have heard of Dota 2, counter strike, league of legends etc. Dota 2 conducts international tournament in which all the teams in the world participate [6]. It happens in Seattle gaming arena each and every year. The prize money is collected from the users itself who buy tickets to watch the event online for free, last year's prize pool for the game was somewhere around tens of millions of dollars and each ticket just costs $7 which shows how big of a user base it has. Online gaming Industry is so vast that people do it as their full time for living, it has become their source of income. People do live telecast of their game play as well every day via several streaming sites like Twitch, YouTube gaming etc. All these tells us the importance of data that gaming industry revolves around and provide us a greater cause to safeguard it from various threats that are present. A study shows that almost all age group people get involved in one or another form of video games which involves playing online with various known and unknown people. ZeniMax media Inc

is a company responsible for analyzing various gaming companies and provide them with information assurance [7]. This company is responsible for several successful stories in the gaming industry. It is always a good idea to obtain helps like this whose main goal is to provide advice and solutions to various security concerns and issues that are present. This is so because even though there could be an internal team of well qualified testers and creators there could be something we can miss as we are all humans. To avoid such situations, it is always a great advantage to seek the help of such huge companies whose primary goal is to find and seal bug in the product like quality testing. The few of the common loopholes or the common security issues that we identified in gaming industry are.

- There is not a proper way of checking attacks like the Man in the Middle attack or eavesdropping etc. The server and client interact whenever there is a information required from the server or there is a need to communicate with another client. There should be some check done by the gaming companies with respect to such medium of communication.
- One of the most recent mess are the ads which the gaming companies post without proper background check which could lead to users being redirected to malicious sites which could steal any kind of information from their personal information to the credit card.
- There is improper encryption in place for data transfers between clients and server-client.

There should be proper measures need to be taken to reduce the number of threats that are present. Every gaming company should shift their importance towards the development of security feature rather than development of new interesting games. A loss or an attack on the games that the company as is equivalent to losing their customer base and their reputation.

B. Wearable Devices.

Wearable devices are the most recent technology advancement. Even though they are the most recent ones the user base and the users involved in it is huge. We do see videos posted on YouTube every day with respect to their amount of run they recorded on Fitbit or the number of calories burnt on their apple watch etc. Wearables does not stop over there but they also get extended in medical field. Wearables are saving lives of millions of people who are having serious medical conditions and cannot survive with the support of technologies. The information that the wearables handle is very sensitive and important as they include huge amount of personal information from location to medical condition.

Wearables can be attacked in several different ways as they are involved in several different environments. Broadly speaking 1) We could either attack the medium of communication between the wearable device and the mobile device. 2) We could attack the medium of mobile device and the server or the wearable and the server and steal huge amount of information. 3) We could also attack the server itself and steal information that is stored in the server and which is used for historical purpose. Even though there are several ways in which a wearable and its information could be compromised the most frequent one that is under attack would be the Bluetooth connectivity of the wearable and the mobile device. The Bluetooth connectivity is under attack very often due to the reason being it is the most fragile one in the world that the Wearables live in. Bluetooth cannot include lots of security features and it is very limited in processing power and battery power which is one of the main reason they are always under attack.

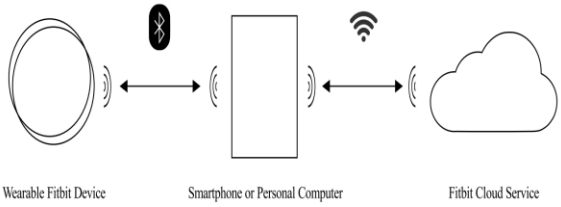

Fig. 1. Werable Device Communication. [8]

Bluetooth is just a standard of wireless communication which is used for communication between devices in a short range. There are several kinds of attacks with respect to Bluetooth which being denial of service (DoS), Man-In-The-Middle-Attack(MITM), eavesdropping, modification of messages etc. It is important that we study the security of Bluetooth in general and provide various solutions that could be used for all these kinds of attack.

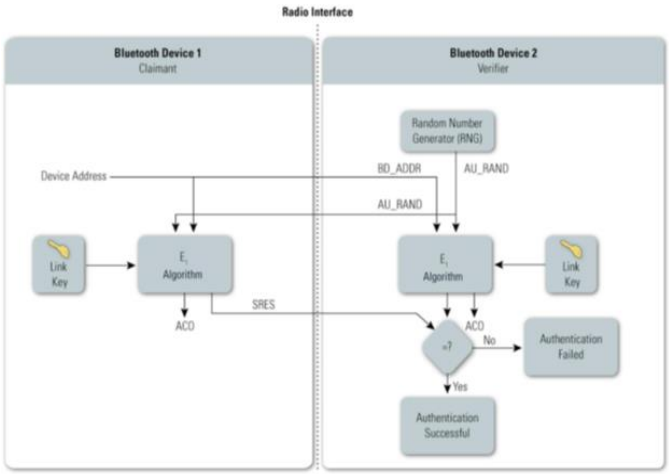

Fig. 2. Authentication Of Bluetooth Device. [9]

There are three broad spectrum of security features provided by the Bluetooth technology, these are the three different modes in which a Bluetooth device can operate. These modes are Mode I, Mode II and Mode III [9]. Mode I involve very little to no security and thus it should never be used when manufacturing the Bluetooth device. In this mode, there is no security feature which is used for authenticating the connectivity thus any device can send any data to any other device and they communicate. This mode also lacks the data encryption mechanism thus when two devices communicate all the information are transferred as native text. The Mode II as medium security features, this mode is also known as the

service level security mode. The mode is called as service level security mode because the security feature is incorporated into the device by the application and not the device itself. Thus, if there is no security with respect to the application which handles the Bluetooth device then there will be no security. This way of handling security we could skip several security threats or just cannot implement counter measures for them in this way. Mode III is the security mode which as the highest level of security feature and this mode should be the one which would be used by all the device. This mode provides security mechanism for both authenticating the devices for connectivity and also for encrypting the data to transfer between the devices. Even though there are these modes available but there always the flexibility or the decision to choose the mode rests in the hands of the manufacturer. This should not be the case and the importance should be given to the security and there should a rule that all devices follow the security Mode III which could help resolve lots of problem related to information theft.

C. Social Network Medium.

Social Network is the new medium of communication between humans rather than in person, one of the main reason behind this is the number of features that it brings to the users table when two of the users who want to stay socially connected are miles apart, the accessibility it provides to the user for connectivity which makes the user to connect to another any time and date given. There could not be a single person in the world who do not own a social networking account. There are several types of social networking platforms that are available like Facebook, twitter, myspace, hangouts, etc. Users who use such social networking medium shares huge amount of information with each other which are very sensitive and important. There are several kinds of attacks that could take place in social networking platform like phishing, Clickjacking, Sybil attacks etc. Users play a very huge role in avoiding several of these threats and it should be the responsibility of each of the user to protect his or her own information.

III. DESCRIPTION OF RESULT

There are several recommendations that should be implanted in each of the three environments that are under study to make it a safer tomorrow.

A. Online Gaming Industry.

It is the duty of both the game developers and the users to safe guard their information and to use the platform in a proper way so that one can avoid information theft.

- Users should be cautious while sharing their information. They should also take extra care on with whom they are sharing their information with.
- Users are advisable to maintain a separate identity from their email address to everything so that the attacker would not be able to easily track them and exploit them with their information.
- Users should be extra cautious and try to avoid sites which are they not aware of for payments etc.
- The companies should train and provide proper ground rules to all the employees of the company so that they can follow them for advanced security measures.
- The main object or goal of the company should be to maximize the security level of the application and handling of information within the company.
- All the games should be carefully monitored, tested and examined in a beta testing environment before deploying it into production.
- There should be proper encryption when it comes to information exchange via a medium with respect to the product.

B. Wearable Devices.

Wearables are very important part of human survival just like the mobile phones for communication. Wearables improve the living standard of several individuals in their day to day life activity from fitness tracking to monitoring their heartbeat. They cannot be eradicated but for the future industry of wearables there should be a proper platform and ground rules deployed so that one can avoid past and future security threats.

- Each of the manufacturer must provide the users with proper awareness about their device and what will be their responsibility to use the device in an effective and hassle free manner.
- Each of the manufacturer must do an internal check on regular basis and improve the security measures that are in place and check whether they are being followed properly.
- There should be proper maintenance of the inventory of the devices that are being sold and which are active and which are not so that it would be easy for tracking and identifying the bug in a future perspective.
- The Bluetooth devices must always use the highest level of security feature that could be incorporated.
- The level of range a Bluetooth device as should be very minimal which avoid the range needed by the hackers for hacking the device there by reducing the hackers as well.
- Security PIN on the Bluetooth device must change often so that to avoid easy hacking of the device.
- The password that is present for protecting the Bluetooth device must not be same as the other ones the user use and must not be easily predictable.
- The data transmitted should always encrypted with high level of encryption.
- There should be possibility of inducing biometric security feature into the device for the extra layer of security.

C. Social Network Medium.

- Users must change their password on a regular fashion.
- Uses must use the two-factor authentication feature if it is present to be able to provide the extra layer of security feature for protecting from the hackers.
- The information shared via social networking sites should not be very sensitive ones.
- Do not get hacked by clickjacking. Make sure you click the proper and known links and use a search engine to access the same link rather than becoming a victim of clickjacking.
- Always add only those people whom you are aware off and ignore unknown profiles.

IV. Contribution To The Project

The project and team was subdivided based on the three main environments of study and each sub team where given with a topic of research. The team that I belonged to was wearables. The goal of our team was to study about the newest trends in wearables and the security issues concerned with the same. We choose a single topic which would be the sub topic of wearables and built upon the project on it. The topic that I chose was Bluetooth Security and I was responsible for researching on it. All the subtopics and information provided under Bluetooth security was a research work of me. I was also responsible for the recommendation part for wearables and I wrote the conclusion and recommendation for wearable devices. I had to research six papers and several online resources to provide my team with solid information regarding the wearables and its connectivity. The topic of Bluetooth security was spoken by me during presentation and the slides regarding Bluetooth security was prepared by me.

V. Knowledge Acquired

The project provided us with vast amount of knowledge regarding information assurance. The project thought us on the importance of Information security, the drawbacks present in current trends and human life with respect to information security. The project provided great insight into security of wearables, the Bluetooth security, the connectivity, the data in cloud security etc. It gave us insight into information assurance on social gaming paradigm. The importance protecting information in social networking etc. It provided us with knowledge on the role of myself as a developer in the developer community and measures that I need to take to protect my software and the data of users involved with my software.

Acknowledgment

We thank professor Stephen S. Yau for giving such an opportunity to do a research project which helped a lot in understanding the importance of information assurance in daily life. We thank teaching assistant Tamalika Mukherjee for helping us in completing the project.

References

[1] Create Virtual Experiences. (n.d.). Retrieved March 23, 2017, from https://www.lindenlab.com/

[2] Action Adventure and Massively Multiplayer Online Games (MMOG). (n.d.). Retrieved March 23, 2017, from http://www.funcom.com/

[3] Davis, S. B. (2009). Protecting Games: A Security Handbook for Game Developers and Publishers. Course Technology PTR.

[4] Hoglund, C., & McGraw, G. (2008). Exploiting online games: cheating massively distributed systems. Boston: Addison Wesley.

[5] How Online Gamblers Unmasked Cheaters. (2009, June 28). Retrieved March 25, 2017, from http://www.cbsnews.com/news/how-online-gamblers-unmasked-cheaters/

[6] Dota 2. Accessed April 12, 2017. http://blog.dota2.com/.

[7] Mohr, S., & Rahman, S. S. (2011). It Security Issues Within the Video Game Industry. International Journal of Computer Science and Information Technology,3(5), 1-16. doi:10.5121/ijcsit.2011.3501

[8] Cyr, B., Horn, W., Miao, D., & Specter, M. (2014). Security Analysis of Wearable Fitness Device (Fitbit). Retrieved from http://courses.csail.mit.edu/6.857/2014/files/17-cyrbritt-webbhorn-specter-dmiao-hacking-fitbit.pdf

[9] Bouhenguel, R., Mahgoub, I., & Ilyas, M. (2008). Bluetooth Security in Wearable Computing Applications. 2008 International Symposium on High Capacity Optical Networks and Enabling Technologies. doi:10.1109/honet.2008.4810232