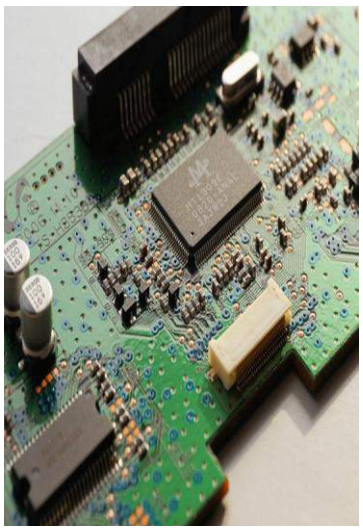
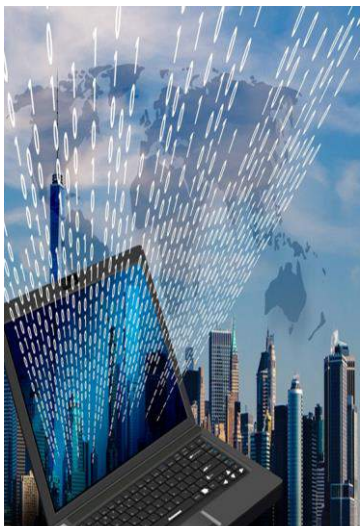
ISBN: 978-0-9957075-2-8

INTERNATIONAL JOURNAL of ENGINEERING AND APPLIED COMPUTER SCIENCE

EMPIRICAL RESEARCH PRESS

Volume: 02 | Issue: 01 | January | 2017



EMPIRICAL RESEARCH PRESS LTD.

Kemp House, 160 City Road, London

United Kingdom

IJEACS

International Journal of

Engineering and Applied Computer Science

Empirical Research Press Ltd.

London, United Kingdom

Volume: 02, Issue: 01

ISBN: 978-0-9957075-2-8

www.ijeacs.com

Indexing, Hosting and Advertising

Internet Archive

Message

International Journal of Engineering and Applied Computer Science (IJEACS) is an open access, double-blind peer reviewed international journal, monthly online publishing by Empirical Research Press Ltd. Empirical Research Press is a research publishing company with name, trademark registered, incorporated in England and Wales, United Kingdom.

The scope of International Journal of Engineering and Applied Computer Science is to publish high quality research contributions, latest innovations, advance development carried out in the field of Engineering, Computer Science and Technology. The original research, review, case study, survey, new interpretation and implementation of concepts and theories, recent technological development, technical reports, empirical discussions are invited to submit for publication.

The major objectives of International Journal of Engineering and Applied Computer Science are to maintain high quality of publications and indexing with world's highly accessed and cited research and academic databases. The scope of IJEACS publications also includes special and interdisciplinary research contributions. We offer best wishes to readers, reviewers and contributors of IJEACS.

Board Members of IJEACS

Content

International Journal of Engineering and Applied Computer Science (IJEACS)

Cyber-Defensive Architecture for Networked Industrial Control Systems

Charles Kim
Electrical Engineering and Computer Science
Howard University
Washington DC, USA

***Abstract*—This paper deals with the inevitable consequence of the convenience and efficiency we benefit from the open, networked control system operation of safety-critical applications: vulnerability to such system from cyber-attacks. Even with numerous metrics and methods for intrusion detection and mitigation strategy, a complete detection and deterrence of internal code flaws and outside cyber-attacks has not been found and would not be found anytime soon. Considering the ever incompleteness of detection and prevention and the impact and consequence of mal-functions of the safety-critical operations caused by cyber incidents, this paper proposes a new computer control system architecture which assures resiliency even under compromised situations. The proposed architecture is centered on diversification of hardware systems and unidirectional communication from the proposed system in alerting suspicious activities to upper layers. This paper details the architectural structure of the proposed cyber defensive computer control system architecture for power substation applications and its validation in lab experimentation and on a cybersecurity testbed.**

Keywords- Component; Supervisory Control and Data Acquisition System, Smart Grid, Power Substation, Cybersecurity, Diversification, Testbed.

I. INTRODUCTION

Cyber incidences are ever increasing as they are expanded from simple bragging intrusion to monetary gains and exploitation to trading secret stealth and to military and national security espionage. One important area in the cyber incidences in which public are not keenly aware of is networked embedded computer systems for intelligent and autonomous control and processing applications including, but not limited to, smart power grid, water treatment and distribution systems, petro-chemical plants and refineries, and mobile and home automation systems, termed combined as Internet of Things (IoT).

The widely adopted IoT on open network architecture provides the benefit of economy of operation; however, unfortunately, it opens the door for unintended threats including malicious code manipulation, data gathering, and unauthorized intrusions into the network. A successful intrusion would allow attacks on operator consoles, and harmful access into control functions which would consequently disrupt normal operations and thus pose a public safety threat.

Presently, the hardening of system is heavily focused on the cyber security for information systems connected to the Internet, and there are numerous strategies and tools available, and are under development. Anomaly and intrusion detection, network access behavior analysis, modeling approach, mitigation are just a few of them. Understanding attack vectors is essential to building effective security mitigation strategies. Attack vectors include viruses, e-mail attachments, Web pages, pop-up windows, instant messages, and deception.

There are several common countermeasures proposed against attack vectors [1]. They include: (i) development and review of security policies; (ii) employment of blocking access to resources and services on the network; (iii) enactment and monitoring of detection of intrusion and malicious activities, (iv) implementation of mitigation against possible attacks, and (v) application of continuous fixing, upgrade, and patch the software vulnerability.

However, the countermeasures developed from metrics can block some attack vectors but are not totally attack-proof. They are backward-looking metrics and measures, analyzing only after an incident with subsequent damage has already occurred. Therefore, the metrics and measures and mitigations developed for the Internet and computer networks may not be effective in dealing with unknown malwares and vulnerabilities specifically targeted for safety- and mission-critical control system applications. The Stuxnet malware attack to an Iranian nuclear facility demonstrates that the reality of the vulnerability of safety-critical systems to cyber-attack is real, and that there will be dire consequences to critical infrastructure if such cyber threats are not detected and mitigated properly and timely [2].

Considering the impact and consequence of mal-functions in the safety-critical applications caused by cyber incidents, this paper proposes an architectural change in the way components are structured so that a networked control system becomes cyber-defensive and resilient even under

ISBN: 978-0-9957075-2-8

compromised situations. The proposed system aims to be insensitive to variations in inputs, processes, and outputs of cyber contents. The proposed defensive architecture is centered on diversification of hardware systems and unidirectional communication to the energy management system for alerting suspicious activities. The rationale of the architectural approach against cyber threats is the plain truth that it is impossible to predict cyber events throughout the computer control system's lifecycle, and that detection and mitigations strategies may be good for old and known malwares and viruses only [3]. Therefore, the methodology used in the proposed cyber-defensive architecture for power substation control systems focuses, assuming that an attack will occur, on designing a system that is robust enough in its response so that the effect of an attack will be minimal and the power substation can continue in serving customers and in protecting power systems [4].

The paper is organized as follows. In the next section, we discuss about the present computer control systems deployed in a power substation and their vulnerabilities. Then, we detail the proposed architectural approach with hardware and software diversity to be insensitive to the cyber inputs and activities, which would results in cyber-robust and cyber-resilient systems. After that, validation of the new architecture is examined on a cybersecurity testbed and in lab experimentation. Then, we conclude the paper.

II. VULNERABILITY OF INDUSTRIAL CONTROL SYSTEMS

Over the past several years, power substation systems have become highly sophisticated in structure and operation, featuring various types of intelligent devices that allow advanced operation and control functions. Computer and communication technologies have transformed stand-alone computerized control systems to Internet-connected smart grid control systems. The smart-grid network provides a great benefit of situation awareness, data collection and analysis for operational efficiency, and coordination of automation and restoration of power networks [5].

A lot of the devices that constitute these smart systems are seen commonly more demanding sectors. Common examples of these devices include smart meters, phasor measurement units, and sensors (voltage and current monitors) and actuators (circuit breaker openers/closers). These "intelligent electronic devices (IEDs)" are networked, as remote terminal units (RTUs) of a supervisory control and data acquisition (SCADA) system, which in turn is connected to an enterprise network or energy management system from which engineers are allowed to operate IEDs and, when necessary, control request or resolve their problems. The advantages afforded by remote access has necessitated the use of Internet and wireless networks, and subsequently, SCADA networks are no longer "air-gapped" but are usually connected to their corporate network and internet through a firewall. This relatively open connectivity has in turn resulted in an increase in security vulnerabilities [6].

To illustrate a sample of the vulnerabilities of the current control and protection system in power substation, a representative diagram is given as Fig. 1. The diagram highlights a simplified representation of a power substation with a communication network (CN) server and a computer/digital relay is disposed for a circuit breaker operation. The enterprise-level energy management system (EMS) is connected to the substation via the Internet. The CN device connects the substation systems to the Internet where all of the engineering staff can login and access the system. The EMS monitors multiple substations via the Internet, and the flexibility of the network allows engineers to control and monitor the relay system from off site.

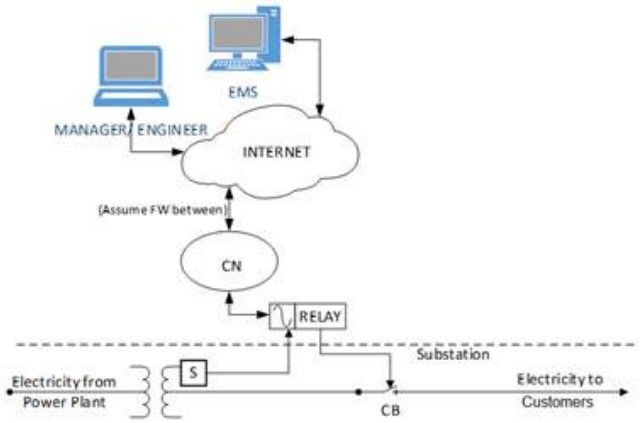


Figure 1. Simplified Representation of Present Power Substation System.

A current sensor (S) is attached to the relay, and based on the sensor reading, the relay can open the circuit breaker (CB) by sending a command signal to the CB actuator when it is necessary or in an emergency. The relay is built on a computer with a standard operating system such as Windows that executes a program that is coded for specific functions and features. When the relay is programmed as an overcurrent protective device, if the sensed current level is higher than a threshold, it would generate an "Open" signal for the CB actuator. It is assumed for our discussion that the relay is an overcurrent computer relay. In addition, a standard desktop computer labeled as 'Manager' with a designated operating system is connected to the relay via the Internet. This allows the individual responsible for overseeing the proper functioning of the system to manage and control the relay, should the need arises.

Now consider cyber vulnerabilities of the substation depicted in Fig. 1. First, the Internet connection represents a possible entry point for hackers to infiltrate the system. If a hacker can gather the appropriate login credentials of the communication network server, he/she can possibly gain access to the relay and alter its operating state. Once that party is logged into this system, they have free reign to enact whatever change they please, which we are assuming is to damage the system in some way. Any alteration to the relay may have major repercussions for the substation and the consumers served by the substation. It would also have a direct effect on surrounding substations as the load of the compromised substation would have to be redistributed amongst its neighbors. This possibility recalls the Federal

ISBN: 978-0-9957075-2-8

Energy Regulatory Commission's finding that the U. S. could suffer a coast-to-coast blackout if just 9 out of the country's 55,000 transmission substations are knocked out on a scorching summer day [7].

III. CYBER-DEFENSIVE ARCHITECTURE

As mentioned above, there exist vulnerabilities in the present power substation and its network, and the countermeasures developed from the presently employed metrics are not attack-proof. Moreover, metrics and measures and mitigations developed for the Internet and computer networks may not be effective in dealing with unknown malwares specifically targeted for safety- and mission-critical control system applications. Even dynamic and learnable measures and metrics cannot possibly detect all and, particularly, unknown and new malwares and their tactics. Therefore there is a demand to make computer control systems robust against cyber threats and resilient under such cyber-attacks.

The proposed architectural approach aims to be cyber-insensitive, and the logic of the proposed defensive architecture is grounded in the concept of software and hardware redundancy/diversity and of utilization of unidirectional network connection. More specifically, the architecture is the result of combining the standard principles of diversified redundant hardware and software for defense-in-depth into a very efficient supplementary system that can integrate with the general structure of the systems currently in use.

A. Diversified Redundancy and Defensive Architecture

The use of redundancy design techniques is already an accepted practice when trying to address fault and failure scenarios in software and hardware. For example, most data is typically backed up to secondary storage spaces and synced as often as possible to ensure minimal to no operational disturbance in most industries. Also, critical manufacturing or generation processes are built with redundant hardware measures to allow easy replacement, repair and maintenance.

Redundancy is effective, but if a machine fails due to a virus attack, for example, then even the redundant machines will be susceptible to the same virus, if they are of the same hardware and software version. This common-cause failure would most likely damage both machines. If, however, the redundant machine has different hardware specifications, there is much greater probability that the redundant machine would survive against the same problem which has caused the primary machine to transition into a fail state. This difference in hardware (and software) is called diversity. Design diversity has also been a tried and true method employed to add a layer of protection to critical systems by protecting redundancy systems from such common-mode failures. Its range of application is vast and its representation can be in the form of software variants to actual physical design differences between primary systems and their redundancy counterparts.

A representative model of the proposed system architecture is illustrated in Fig. 2. In the proposed design, alongside the existing primary computer/digital relay ("RELAY"), there is secondary digital relay that functions, in sensing the current level and generating a signal for CB operation, identical to the existing one but built on different hardware such as field programmable gate array (FPGA) and run on a completely different software environment ("FPGA"). Unlike in the existing system, the CB operation signals from the two relays are monitored and selected by a supervising computer system ("SUPERVISOR") which is built on a PC or a hard-wire system; therefore, the SUPERVISOR is in charge of the eventual control of the CB. As in the existing substation, the primary RELAY is connected to the communications network (CN), while the secondary FPGA is remained not connected to any network.

The SUPERVISOR is separated from the Communication Network, and reads the CB control signal outputs of both relays and decides if either one is erroneous or not by conferring with a database server which contains data readings collected at the sensors and the corresponding CB operations over an extended period of operational hours. Under regular operating conditions, there should be near perfect correlation for given sensed value between the CB control signal generated by the two relays and the cached CB operational mode in the database server.

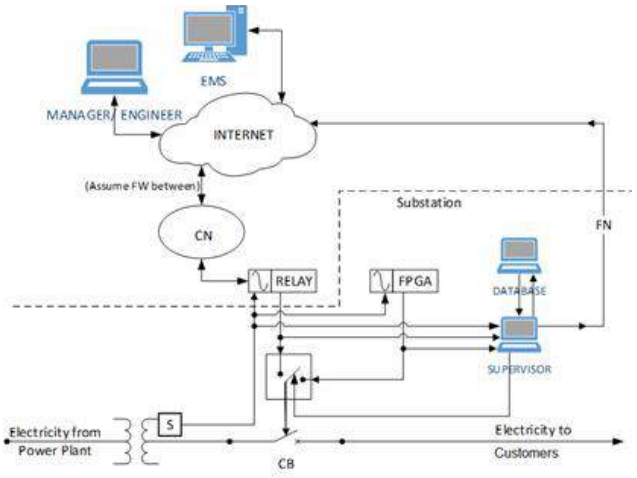


Figure 2. Defensive Architecture for Power Substation System.

In the event that the SUPERVISOR finds an inconsistency between the CB signal of the primary RELAY and the database, for example, it gives the CB operational control to the secondary FPGA relay which produces the correct signal. At the same time, it sends a warning message to the EMS via a unidirectional fiber network (FN) as this is indicative of the primary RELAY being possibly compromised, to alert the management personnel of the state of the system. The importance and distinct advantage of using unidirectional network connection is the fact that this new system at no point is required to receive and act on requests [10]. Hence, the integrity of alerting is preserved and the possibility of communication related intrusions such as Denial of Service (DoS) attacks is inherently prohibited.

ISBN: 978-0-9957075-2-8

B. Qualitative Assessment

Before we validate the new approach, let's do qualitative assessment of the new architecture on its claimed strength against cyber-attacks under a few instances, all feasible within the environment of power substation operations. First, we consider the presence of a common computer virus which gains entry into the system via a negligent substation engineer. Under these circumstances, the operation of the primary RELAY becomes compromised. The secondary FPGA relay on the other hand, by way of design diversity remains unaffected. Even in the instance of viruses that have the ability propagate across networks, the difference in programming methodology between the two relays grants mutual exclusion in the case of software attacks, eliminating the threat of common-cause virus infections.

The next attack scenario considered is a theoretically attempted man-in-the-middle attack. This scenario involves attacks in which access credentials are mined from unsuspecting parties. In this case, it is difficult to determine if the system is under attack because the information used to gain unauthorized access to the system is indeed legitimate. Therefore, changes can be made to the primary RELAY without any intrusion indicators being set off. But even in this highly compromised state, by virtue of the comparison check that occurs continuously at the SUPERVISOR, any changes or discrepancies generated by the intruder are flagged, and controlling of the substation functions is committed to the secondary FPGA relay.

Another considered scenario is an event of common hardware and software failures. Hardware failures in this context refer to incidents such as purposeful or accidental physical damage and hardware component faults, which cause eventual failure. The proposed architecture ensures that in case of damage to the primary RELAY, the secondary FPGA relay can function autonomously not being susceptible to common-mode hardware failure. This ensures that the service is maintained until the proper repair and replacement procedures can be carried out. While the probability of simultaneous failure of both the primary and secondary relays of the proposed system exists, it is theoretical and very small.

The last scenario we consider for qualitative assessment is with a mode of attack employed by a Stuxnet-like worm in its various iterations. The Stuxnet worm is a program that was developed to target specific industrial software on a specific brand of equipment in a plant [2]. This type of specialized attack is hard to defend from because it relies on targeting and exploiting certain vulnerabilities in the operating system. Fortunately, the design diversity afforded by the new system structure acts as a functional safeguard. Having both relays run on very different software and hardware architectures ensures that whatever damage is done is limited to the primary RELAY. The new solution architecture upgrades the existing system to a multi-tiered, cooperative system in which desired relaying functions are kept intact all the time fulfilling the response robustness required of systems that handle such critical task as power substation control and management.

To verify the feasibility and viability of the proposed architectural solution, two approaches are employed: simulation on a cybersecurity testbed and experimentation with hardware components. In both validation approaches, the representative models of the existing power substation system and the new defensive-architecture system are implemented in a network simulator and in a system of microcontrollers and FPGA, respectively. Then, the two models are subjected to the same attack conditions and the each response is recorded and analyzed. The testbed simulation is discussed in the next section, and the hardware lab experimentation follows in the section after the next.

IV. VALIDATION IN TESTBED EXPERIMENTATION

To accomplish verification via software simulation, a testbed known as DeterLab (cyber DEfense Technology Experimental Research Laboratory) is utilized. DeterLab is a facility for scientists engaged in new cybersecurity technologies. The Deter Team works with subject matter experts in specific areas of cybersecurity or critical infrastructure protection, and the DeterLab is a part of the work which provides real world capability to research, develop, discover, experiment on and test cyber defense technologies [9]. Approved users can access DeterLab's advanced resources and tools, and perform repeated, verifiable experiments. DeterLab provides over 400 computer nodes, with up to 10 network interfaces per node, each of which can support multiple apparatus elements by using virtualization techniques that support the experimenters' goals [10].

A. Existing Control System Experimentation

To demonstrate the vulnerabilities in the existing system of Fig. 1, we model the network topology as Fig. 3. The DeterLab evaluation begins with conceptualizing the model of the simplified primary Relay. The model is then created in DeterLab after which a user interface is created in DeterLab. A remote administrative tool (RAT) is used to show how an intruder can infiltrate the system and change Relay configuration files on the EMS. The RAT is classified as a virus called Trojan horse program, a malware which by itself is not capable of automatically spreading to other systems. Trojans are usually downloaded from the Internet and installed by unsuspecting users. They typically carry payloads or other malicious actions that range from the mildly annoying to the irreparably destructive. They may modify system settings to start automatically [11]. As shown in Fig. 3, the model consists of an EMS, Intruder and Engineer nodes connected to the internet. The primary, networked Relay is connected to the internet through a Router (or firewall not shown).

In the topology of Fig. 3, since this is a virtual environment, all nodes are reserved with Class-A IP addresses. Although, in the DeterLab representation, each device in the substation has an IP address, only the EMS is part of the TCP/IP network, with the address 10.1.1.2. The Sensor and the Circuit breaker, not shown, are physically connected to the Relay without communicating directly to the Router.

ISBN: 978-0-9957075-2-8

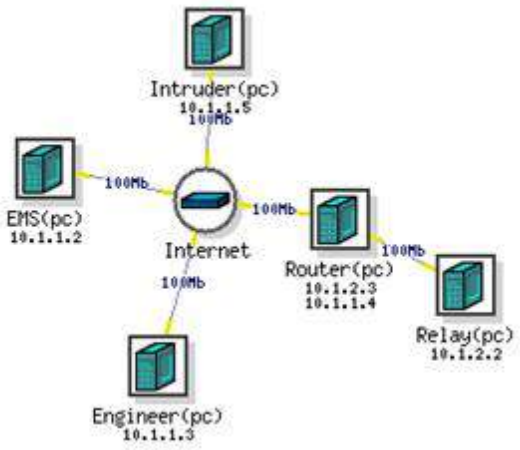


Figure 3. DeterLab Reprentation of the Existing Control System.

Other notable hosts include the Intruder node with an IP address of 10.1.1.5 and the Engineer with and address of 10.1.1.3. The Intruder and Engineer nodes are added to represent a hacker and an engineer in order to stage an attack, respectively. The Intruder in particular is added to simulate an ill-intentioned individual who wants to gain unauthorized control over the system. In order to do this, the Intruder may use a RAT. There are various types of RATs available, and most of them are made to be used with the consent from the owner of the controlled computer, but for our purpose, we will assume no permission is granted. Although various RATs are used for a range of purposes, their structure remains the same: it has both a client (installed on the intruder's machine) and a server module (installed on the victim's machine). The server has a process that initializes as the system boots up and keeps running in the background, waiting for the client to connect. When the intruder wants to remotely manage or control the server, he just launches the client on his machine. If the remote computer is powered up and connected to the Internet, then controlling is possible.

This DeterLab simulation considers the situation where a negligent engineer or manager, who has the authority to access the EMS computer remotely, executes a program that covertly installs the server module of the RAT on his computer. Although most antiviruses easily flag these kinds of files, there are some techniques to make them fully undetectable. However, we will not cover them in this article. Instead, let us suppose the user has no updated anti-virus software installed on his machine and/or the hacker has encrypted his server on purpose.

By using PuTTY to create a tunnel and forwarding the desired ports (local 6789 to remote 3389), it is possible to connect to the nodes at DeterLab using remote desktop connection on Windows. In order to gain this control, the hacker must force the user to execute a file that will automatically install the server module. He can do it by using techniques known as social engineering, that is, by persuading the user to download and execute the file.

Considering that the RAT is a Trojan which can be disguised as a legitimate software component, if the Trojan's configuration complies with the substation network, it will silently enable the remote control, without the victim's (engineer or manager) consent or acknowledgment. On whichever machine the file is run, either the EMS or the Manager/Engineer's computer, it will install the server module. Once connected, the hacker can easily modify any file in the remote system. The RAT used in this network gives the user the possibility to download, modify and upload a file on the host computer. In other words, in power substation, software for controlling the Relay and for activating CB can be modified without the operator's noticing.

To demonstrate how the modification of the threshold value for CB opening by attacks changes the CB signal from the Relay, a user interface is developed using Visual Basic on the EMS. For the interface, since the current sensor cannot be included in the network topology, the reading from the sensor is entered manually, and the CB operational threshold ("HIGH" as illustrated of the interface in Fig. 4) is stored in a file on the primary Relay. Under a compromised situation via RAT attack, the setting for HIGH may be changed from 200 to 150, for example. Then, even under the normal current level of 180 at which the circuit breaker normally remains closed allowing power supply to the customer, the Relay sends out OPEN signal to the CB by the altered threshold of HIGH to150 from 200.

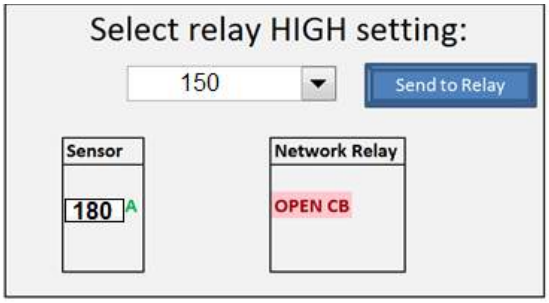


Figure 4. Response of the Network Relay by the RAT attack.

The illustration made above is a common type of cyber-attack and currently infects thousands of computers around the world. A hacker can easily adapt this tool to infect IEDs from someone who possesses high privileges in an electric power company. The executable file can be encrypted or bended to other file or spread through other known means in order to reach its final target. Therefore, it is important not to depend on a single machine to control critical devices like digital protective relays, neither is it recommended to trust people with low awareness on the critical importance of cyber-security to operate this kind of equipment.

B. Defesnive-Architecture Experimentation

The DeterLab simulation for the proposed architecture is performed similarly. First, a model is conceptualized for the new architecture which includes a non-networked diversified redundant FPGA relay and SUPERVISOR. Second, a slightly different user interface is developed to

display the response of the new system for primary Relay and secondary FPGA under the same RAT attack and relay threshold setting file modification. Third, it is shown how a message is sent from the SUPERVISOR to the EMS and how unidirectional flow of data is achieved between them.

Fig. 5 illustrates the topology model of the proposed diversified architecture in the DeterLab. The network topology is created using six nodes. The EMS is connected to the Internet. The primary Relay (REL) is connected to the Internet via a router or firewall (FW1). The other network in the model is the supervisor network (SN) which comprises of the SUPERVISOR (SUP) and database computer (DB). The sensor and the secondary FPGA relay are non-networked devices and thus are not represented in the topology, and their values and responses are simulated by manual entry into the user interface.

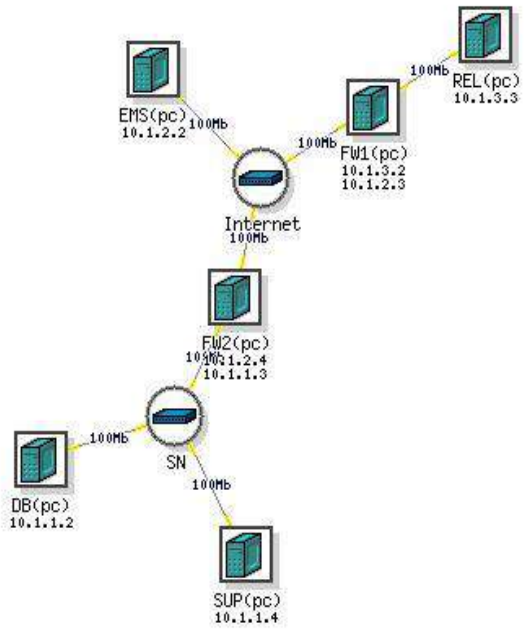


Figure 5. DeterLab Representation of the Defensive Architecture.

The same user interface developed for the existing system simulation is used with a slight revision to accommodate the secondary relay FPGA as illustrated in Fig. 6.

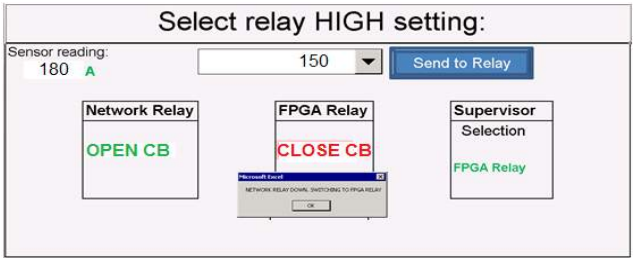


Figure 6. User Interface Display under an attack with modifed setting.

The FPGA Relay provides the same functionality as the primary network relay but with strong immunity against setting modification due to its hard-code environment. If the relay setting is changed by the RAT attack with altered threshold value as done before, from 200 to 150, for example, for the normal loading condition of 180 A, the network relay's output ("OPEN") does not match with the true output ("CLOSE"), and the selection of relays to control the circuit breaker will be switched to the FPGA relay with a message of such control transfer. Fig. 6 shows the switch to FPGA with a message in the user interface.

To demonstrate message delivery from SUPERVISOR to EMS of the abnormal and suspicious behavior of the primary network Relay, a socket program in C programming language is used [12]. The program is written in two parts: a server (fileserv.exe) and a client (fileclient.exe). The server accepts a connection from client through a specific port, receives the file name, creates file with the given file name, receives the file contents, and writes the contents to file. On the other hand, the client connects to the server, sends the file name, and sends the file contents. For the DeterLab simulation, the server module is executed on the EMS while the client module is executed on the SUPERVISOR.

Before executing the program, however, the fileserv.exe file is saved in a folder where the file is to be executed and the message is to be created and saved. Similarly, the fileclient.exe file is saved in a folder where the file is to be executed and the message is to be copied. In both cases, a directory called c:\reports is created. To execute the program at the server a DOS command prompt window is opened and the directory is changed to the c:\reports folder. On the server, the command "fileserv.exe <port number>" is entered. In this case port number 8907 is used. On the client, the command "fileclient.exe" <IP address of destination computer> <port number> <file name> is entered. In this case, the IP address is 10.1.2.2, 8907 is the port number, and NetworkRelayDown.txt is the file name.

Now the last subject of discussion is the unidirectional information transfer from the SUPERVISOR to EMS on reporting the abnormal and suspicious functioning of the primary Network Relay. For our simulation, the unidirectional flow from the SUPERVISOR is set up by using Windows Firewall to block all the ports except for port 8907, the port that is used to send the message to the EMS. All other ports and applications are blocked. The effect of the Firewall configuration is verified in that a computer is unable to connect to SUPERVISOR and unable to ping SUPERVISOR.

C. Discussion on DeterLab Experimentation

Modeling of the existing and the proposed substation systems are realized in DeterLab environment as network topologies with corresponding nodes with the Internet and routers and firewalls. It is shown that with an RAT, it is possible to gain control of a remote computer and change

ISBN: 978-0-9957075-2-8

files on a remote computer, and the existing system sends out an erroneous command to the circuit breaker. On the other hand, the proposed architecture system demonstrates how the secondary FPGA is immune to the attack and the system itself keeps the normal operation mode by the SUPERVISOR's monitoring while the primary network relay, impacted by the changed setting, produces a wrong command to the circuit breaker. In addition, the message is successfully sent from the SUPERVISOR to the EMS using a socket program through TCP. This is achieved by manually triggering the program and setting the port number through which the message is sent. Unidirectional flow from the SUPERVISOR to the EMS is partially achieved using Windows Firewall blocking all other ports except for port 8907. However, the software is limited in that there is no capability of controlling inbound traffic through that port, which means that communication from the EMS to SUPEERVISOR may still be possible if the hacker discovers that port 8907 is open. Also, DeterLab seems to have a limitation in that when the SUPERVISOR is blocked using Windows Firewall, it is not possible to connect to it to do further testing. The experiment would have to be reset by swapping it out and then swapping it in again.

V. Validation in Lab Experimentation

This section discusses a small-scale hardware experimentation of the existing and the proposed new system. The aim here though is not to produce a physically scaled down replica, but to perform an extended version of tangible, logical validation and illustration. It should therefore be noted that the components used to achieve this hardware experimentation are neither directly relatable to the industry specific equipment in use, nor are they scalable. Specific details such as response times are not considered because they would largely be dependent on the precise equipment that would be used if this solution approach is adopted.

A. Lab Experimentation Setup

As for hardware components for the simplified substation systems, as illustrated in the schematic of Fig. 7, the primary relay is represented by an Arduino microcontroller [13] ("Primary Arduino"), the secondary FPGA relay by a Nexys II Spartan-3E FPGA board ("Secondary FPGA"), and the supervisor by an Arduino microcontroller with an attached Ethernet Shield ("Supervisor Arduino").

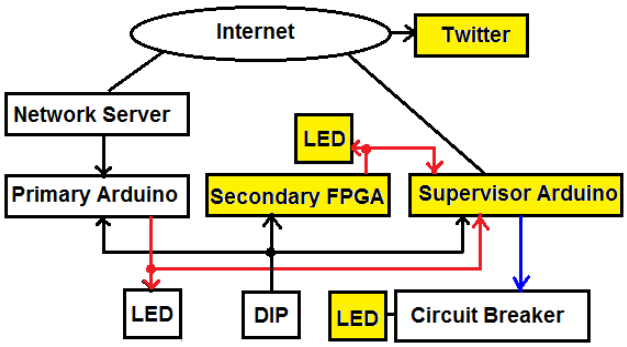


Figure 7. The Hardware Experimentation Components.

The communication server is represented with a laptop with Microsoft Windows 7 Professional operating system which is connected to the Internet. The Supervisor Arduino is also connected to the Internet, and a Twitter account, ArduinoHU, is made to represent the EMS and to simulate the message transmission upon a cyber-incident.

The current sensor is represented by a DIP switch by the position of each toggle of which can simulate various loading conditions. The circuit breaker is implemented by a simple magnetic switch/relay, the operation (Open/Close) of which is controlled by a digital command. An LED is attached to the magnetic relay to indicate the operation state of Open (ON) or Close (OFF). In addition to the LED attached to a circuit breaker representative, an LED is connected to each of the relay representatives to indicate the output status of it. The DIP is directly connected to an input port of both Primary Arduino and Secondary FPGA as well as to the Supervisor Arduino's input port. The digital command to operate the magnetic switch is issued by the Supervisor Arduino from the outputs of Primary Arduino and Secondary FPGA, which are directly connected to the input ports of the Supervisor Arduino.

As for software, a simple code is programmed for the relay representatives for reading values from the DIP switch and sending out corresponding outputs based on the pre-set threshold value. The Supervisor Arduino is coded to take in two outputs and compare them to a database of past sensor readings and respective CB operations which is nothing but a simple table embedded in the code. Fig. 8 depicts the lab hardware experimentation set on a breadboard.

ISBN: 978-0-9957075-2-8

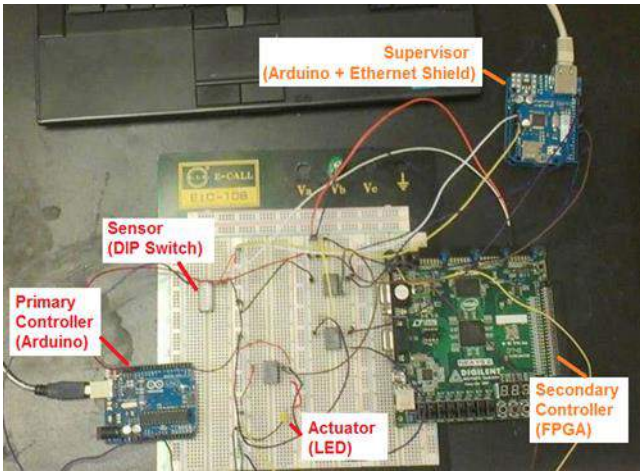


Figure 8. The Lab Hardware Experimentation Setup.

The simulation of the existing system is done on the experiment setup just by using only the network server and the Primary Arduino and the DIP switch (all unshaded components of Fig. 7), and that for the proposed system is done using all the components. In both cases, it is assumed that the attack is made through virtual private network (VPN) of the substation and that the attacker has access to the engineer's laptop after obtaining the credentials from the Trojan virus or using a key logger.

B. Cyber Attack on Existing System

Once the hacker has the credentials to the engineer's laptop which is connected to the substation network, the hacker easily connects, using the remote desktop tool in Microsoft Windows, to the remote communication network server which is also running a Microsoft Windows operating system. In the process, the hacker inputs the IP address of the remote machine and then he types in the credentials for the communication server, to which all IEDs within the substation including the relay (represented as the Primary Arduino in the setup) are connected. At this point the hacker has now access the application that programs the Primary Arduino, and can upload corrupted code to the Primary Arduino.

In this particular simulation, the hacker uploads a code which frequently changes the threshold value for circuit breaker operation from very low to very high, and it results in producing the constantly tripping and closing signals to the circuit breaker, manifested in the blinking LED every one second while the DIP switch positions are remained intact. Under this type of operation, the existing substation system components cannot survive and the service would be disrupted to the customers until the crew come to the substation and repair the problem and restore the service.

C. Cyber Attack on the Defensive Architecture

Now the same attack is staged for the system of the proposed architecture. Again, the amount of loading is simulated with the positions of the DIP switch and a certain threshold value is coded in to Primary Arduino and Secondary FPGA. Also, we assume that the hacker has already entered the network and placed the same corrupted code in Primary Arduino. Since the Secondary FPGA is not connected to the network and keeps its operational logic in its hare-wired code, it does not suffer from the attack. Therefore, while the Primary Arduino produces and sends erratic ever-changing outputs to the Supervisor Arduino, manifesting with flashing LED of its own, the Secondary FPGA sends consistent output based on the loading level. Now the Supervisor Arduino compares the two outputs against the normal operation history from its database, and selects the Secondary FPGA to control the circuit breaker, manifesting the state of the circuit breaker LED the same as that of LED of the Secondary FPGA.

Hence, even under the compromised situation in the Primary Arduino, the intended functions at the substation would survive and there would be no disruption of service to customers. At the same time the Supervisor Arduino sends a twitter message to a Twitter account, ArduinoHU, stating that the Primary Arduino has malfunctioned and alerting the engineers of the EMS to come on site to repair the compromised component. The proposed architecture has shown that it can survive cyber-attacks.

D. Discussions on the Lab Hardware Experimentation

The lab hardware experimentations with the remote attack scenario demonstrate the vulnerability of the existing system and a greater potential of the proposed architecture in surviving cyber-attacks. A minor problem is noticed in simulating the unidirectional message alert from the SUPERVISOR to the EMS via Twitter message. Under this setup and scenario, the message being sent to the Twitter account may be captured and replaced with false message. Even though the false message would not warn the substation system's operation, there is a great chance that no one would be alerted to come to the substation to address the problem. It is hoped that, in real application of the proposed architecture, the suggested unidirectional fiber optic network would do the intended function properly.

VI. CONCLUSIONS

The current period may be appropriately called a cyber-age which has changed every aspect of business operations, factory manufacturing, process operations, and our daily lives in to digital data and cyber bits. The inevitable side effect of this transforming convenience of cyber-age is the cyber threats and attacks whose skills and tactics and targets are not static but constantly evolving. Even with numerous countermeasures supported by government and industry agencies and experts, new threats seem to materialize as soon as old ones are solved or patched. Considering the impact and consequence of the service interruption in a safety-critical application, power grid substation in particular, caused by cyber incidents, a new defensive-architecture based control system is proposed, with expectation that this new defensive architecture would make a networked computer control system cyber-strong and resilient even under compromised situations. The defensive architecture is centered around the diversified redundancy principle and supervised operation with unidirectional

ISBN: 978-0-9957075-2-8

communication against malware attacks. The architectural details of the new proposed system are described along with its advantage in surviving cyber-attacks and in overcoming the vulnerability of the existing system. Also detailed is the evaluation process in DeterLab test bed simulation and the lab hardware experimentation, which demonstrates the validity and survival potential of the proposed defensive architecture system under cyber-attacks.

REFERENCES

[1] Recommended Practice: Improving Industrial Control Systems Cybersecurity with Defense-in-Depth Strategies, October 2009, Department of Homeland Security.

[2] D. Kushner, "The Real Story of Stuxnet," IEEE Spectrum, [Online]. Available: http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet.

[3] R. Langner, Robust Control System Networks, NY, NY: Momentum Press, 2012.

[4] Charles Kim, "Whitepaper – A Diversified Architecture for Robust and Cyber-Strong Control System," Howard University, 2014. [Internal document.]

[5] "What is the Smart Grid?," U.S. Department of Energy, [Online]. Available: https://www.smartgrid.gov/the_smart_grid.

[6] K. Stouffer, J. Falco and K. Scarfone, "Guide to Industrial Control Systems (ICS) Security," May 2013. [Online]. Available: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r1.pdf.

[7] "U. S. Risks National Blackout From Small-Scale Attack," The Wall Street Journal, March 12, 2014. [Online]

[8] R. HA'AYIN, "Pike Research: Unidirectional Gateways Among Most Promising SCADA Security Technologies," PR News Wire, 13 Sep 2011. [Online]. Available: http://www.prnewswire.com/news-releases/pike-research-unidirectional-gateways-among-most-promising-scada-security-technologies-129710343.html.

[9] DETER Team, "The Deter Project," [Online]. Available: http://www.deter-project.org/..

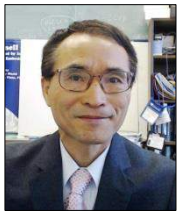
[10] DeterLab, "Projects that have actively used isi.deterlab.net," USC Information Sciences and University of Utah, 2012. [Online]. Available: https://www.isi.deterlab.net/projectlist.php.

[11] "Trojan-Threat Encyclopedia," Trend Micro USA, November 20, 2014 [Online].

[12] B. Mitchell, "Sockets and sockets - introduction to sockets," About.com, [Online]. Available: http://compnetworking.about.com /od/itinformationtechnology/l/aa083100a.htm..

[13] "Arduino," [Online]. Available: http://www.arduino.cc/.

AUTHOR PROFILE



Charles Kim is a professor in Electrical Engineering and Computer Science at Howard University, USA. He received Ph.D. degree in Electrical Engineering from Texas A&M University, USA in the year 1989. Prof. Kim's research includes application of physcis of failure to aero, naval, and ground systems of electrical and electronic devices and networks. Also, he has worked for safety and security for safety-critical systems in automotive and energy industrues. Prof. Kim is a senior member of IEEE.

ISBN: 978-0-9957075-2-8

Computer Aided Development of Fuzzy, Neural and Neuro-Fuzzy Systems

Priti Srinivas Sajja
Department of Computer Science and Technology
Sardar Patel University
Vallabh Vidyanagar, India

***Abstract*— Development of an expert system is difficult because of two challenges involve in it. The first one is the expert system itself is high level system and deals with knowledge, which make is difficult to handle. Second, the systems development is more art and less science; hence there are little guidelines available about the development. This paper describes computer aided development of intelligent systems using modem artificial intelligence technology. The paper illustrates a design of a reusable generic framework to support friendly development of fuzzy, neural network and hybrid systems such as neuro-fuzzy system. The reusable component libraries for fuzzy logic based systems, neural network based system and hybrid system such as neuro-fuzzy system are developed and accommodated in this framework. The paper demonstrates code snippets, interface screens and class libraries overview with necessary technical details.**

Keywords: Fuzzy logic, Neural network, Neuro-fuzzy systems, Soft computing, Automatic development.

I. INTRODUCTION

Advancement of science and technology has been increasingly utilized as a major tool for uplift of mankind. Innovations of modern information and communication technologies (ICT) made human life smoother and problem solving has become easier. It is observed that usage of modern ICT (like Internet) has become ambient and ubiquitous. In spite of availability of lots of tools and technologies, expectations from mighty machines are continuously increasing and demand for more and more human like intelligent systems in various fields has evolved.

Development of an intelligent system is a challenging job due to several reasons. The prime among them are abstract nature of knowledge, volume of knowledge, lack of knowledge acquisition and representation techniques and lack of models/quality standards for the development of an intelligent system. Typical Artificial Intelligence (AI) techniques facilitate development of intelligent system with the aforementioned problems. However, they lack self learning, human like interaction, processing and require lot of efforts as well as cost. There are some new AI techniques such as bio-inspired techniques that offer some advantages above the typical AI techniques. These techniques include artificial neural network, swarm intelligence, fuzzy logic etc. and their hybridization (such as neuro-fuzzy approach). These techniques are sometimes also known as soft computing techniques. Many tools are available to develop such soft AI based intelligent systems. However, these tools are costly, less user friendly and application specific. Further, these tools need training and practice to certify their efficient utilization. There is a requirement of a generic tool which interacts through native/natural language of non-computer professionals/users, reduces effort of development of an intelligent system and saves time of development. No such generic tool is available as per the literature found at present. This paper presents a design and implementation of generic tool that facilitates automatic development of soft computing intelligent systems such as neural network based systems, fuzzy logic based systems and hybrid neuro-fuzzy systems in a given domain.

The paper organization is as follows. Section 2 of the paper presents national and international scenario related to the aforementioned problem and discusses limitations of the existing solutions. The aims and objectives of the proposed research work are enlisted in this section.

Section 3 describes architecture of the proposed framework which can be utilized as generic and user friendly tool that facilitates automatic development of soft computing intelligent systems. An important component of the framework is library of reusable components to develop intelligent systems. Design and implementation issues of the centralized library along with other necessary components are described here. At the end, the paper concludes with advantages and application of the framework generated in different other areas. The section also presents future scope of the research work.

II. LITERATURE SURVEY

To develop soft computing system techniques such as artificial neural network (ANN), type 1 and type 2 fuzzy logic (T1FL and T2FL) and hybrid neuro-fuzzy systems are vastly used. This section presents in brief about each techniques and work done in different areas using the respective techniques.

A. Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) is an interconnected group of artificial neurons that uses a computational model for processing information based on a connectionist approach to computation to impart self learning and other characteristics associated with intelligence. ANN is powerful tool for modeling intelligent systems when self learning is most desired feature and generalized knowledge is not available or can not be documented easily. ANN has ability to identify and learn the correlation between the input patterns, data or cases provided to it. Due to this feature of ANN, it is increasingly being used in intelligent systems where one has little or incomplete understanding of the problem under research but experience or training data of domain expert is readily available. The popular models for implementing ANN are Hopfield model, multilayer perceptron with supervised learning, Kohenon model (self organizing map), and recurrent network. Programming environment like java, c# and other third generation programming languages help in implementing ANN based systems. There are some packages like matlab (www.mathworks.com/products/matlab) are also available which facilitates development of such system.

B. Fuzzy Logic

Fuzzy logic is a multi-valued logic conceived by Zadeh [1] and used to achieve human like loose categorization of objects into classes without boundaries. The fuzzy logic is based on fuzzy sets. Every given object has the partial (fuzzy) belongingness of the concerned fuzzy sets, which is determined by its membership function. For example, if the temperature is 18 degree centigrade, then it belongs fully (say 1) to the set of comfortable temperature. If the temperature is 12 degree centigrade, then it belongs partially (say 0.4) to the set of comfortable temperature. Young lady, rich person, luxury car and comfortable temperature are a few examples of linguistic vague words which can be interpreted by machine with the help of fuzzy membership functions, hence can be easily used in 'if…then…else' rules. System with mainly such fuzzy 'if…then…else' rules are capable of representing knowledge and are known as fuzzy rule based system. Mamdani [2] and Takagi & Sugeno [3] are vastly utilized basic modeling techniques for fuzzy logic based systems. Further, Jerry Mendel has proposed an idea of type 2 fuzzy logic which was an extension of the original fuzzy logic proposed by Prof. Zadeh [1]. The basis of type 2 fuzzy logic is type 2 fuzzy sets. Type 2 fuzzy sets incorporate uncertainty as extra third dimension which gives much clear and logical information about the problem under research. Further, type 2 fuzzy systems need to accompany with type reducer component that converts type 2 membership functions into simple fuzzy membership functions. The popular tools to develop a fuzzy logic based system are third generation programming languages such as java and c# of .Net framework (www.microsoft.com/net) and packages like Matlab (www.mathworks.in). Many specific packages also have been developed such as Fuzzy Attitude (www.fuzzytech.com), JFuzzyLogic (sourceforge.net/projects/javafuzzyed/), Fuzzy Editor (http://jeux.windows.simplenet.com/), etc.

C. Hybrid Systems

Every technique has its advantages and limitations. The idea behind hybridization of two or more techniques is to get advantages of all candidate technique in a common application. For example, consider hybridization of ANN and FL. It is observed that both ANN and FL have their own pros and cons. ANN based systems are good where there is availability of data but lack of generalized knowledge behind it. ANN systems are good in self learning however lacks in documentation of knowledge. This is the prime reason why such system cannot provide reasoning and detail explanation of decision made. Fuzzy logic is good in handling uncertainty and handling natural linguistic values, but lacks self learning and enforces documentation of knowledge in generalized form. Hybridizing these two technologies provide dual advantages of FL and ANN both in one common application. Specifically Neuro-Fuzzy hybridization (NF) achieves advantages of self learning, explanation and reasoning and user friendly interface along with the advantages associated with documentation of knowledge. As stated earlier, third generation programming languages and packages like Matlab can be used to develop such hybrid systems.

Fuzzy Adaptive Learning Control Network (FALCON) [4] Adaptive Neuro Fuzzy Inference System (ANFIS) [5], Generalized Approximated Reasoning based Intelligent Control (GARIC) [6], Neuro-Fuzzy Control (NEFCON) [7], Fuzzy Inference and Neural Network in Fuzzy Inference Software (FINEST) [8], Fuzzy Net (FUN) [9], Evolving Fuzzy Neural Network (EFuNN) [10], Self Constructing Neural Fuzzy Inference Network (SONFIN) [11], etc are popular tools that help in development of such hybrid system.

Not only NF system, but also different techniques such as genetic algorithm and fuzzy logic, neuro-fuzzy-genetic and neuro-genetic hybridization are also popular. At present our framework was designed and implemented to support development of FL, ANN and NF type of systems. However, it can be enhanced to encompass other soft computing techniques and tools. Hence, the literature survey is restricted to these approaches only.

The above tools are either costly or application specific such as Cihan H et al. [12]; or not web-enabled. Many of them are not user friendly and to use such tools would be challenge for the non-computer professionals. Tools like Adaptive Network Fuzzy Inference System (ANFIS) and Dynamic Evolving Neuro-Fuzzy Inference System (DENFIS) need platform of MatLab which is costly. Many researchers have experimented development of dedicated applications in the field. Pioneer of them can be given as Ajit Abraham [13] Jang et al. [14], Mendel [15] and Wu & Mendal [16], Emilio Soria-Olivas et al. [17], Ching Long Su et al. [18], John & Coupland [19], Oscar Castillo & Patricia Melin [20]. Much application specific work is done by various researchers including Malkawi & Murad [21], Nie et al. [22], Bouzaidaa et al. [23] and Azriyenni & Mustafa [24].

ISBN: 978-0-9957075-2-8

Considering the aforementioned work; following observations can be made.

- Majority of the existing solutions are application specific;
- The solutions may not be web based;
- The solution, which supports computer aided development may not support modern artificial intelligent techniques;
- The existing solutions may not be flexible and extendible to accommodate users new requirements and other technology in future;
- The exiting solutions/tools may not be reusable; and
- The exiting solutions are generally meant for programmers hence they may not be user-friendly to non-computer professionals; etc.

This leads to a development of generic, web-enabled and user friendly architecture that supports interactive development of all type of soft computing intelligent system. As a prototype, the framework is designed and implemented for automatic development of FL, ANN and NF hybridized systems. However, it is designed in a flexible way to support many more latest technologies. The prime objectives were decided as follows.

- Development of fuzzy logic based editor to facilitate working with linguistic variable and vague input;
- Development of popular fuzzy membership functions with fuzzification and defuzzification techniques;
- Development of type reducer component for converting type 2 fuzzy systems into typical fuzzy systems so as the above fuzzy function can be used;
- Development of reusable library containing activation functions and learning algorithms for artificial neural network;
- Mechanism for Neuro-fuzzy hybridization; and
- Development of miscellaneous features such as back up, security feature, web templates and theme library of variety of presentation of generated code, etc;

Section 3 describes the design of the framework.

III. Design Of The Framework

Design of the framework which meets objectives finalized and mentioned in earlier section is presented here. The framework is divided into three layers. In its first layer repositories for reusable codes are stored as generic independent objects. The reusable codes included here are neural network, fuzzy logic and neuro-fuzzy systems. Second layer is a database layer. The database accommodates third party tool, user profiles, meta data repositories and protocols, if any. Third layer is an interface layer. The interface layer accommodates information acquisition and interaction facilities for users of the system. This layer also provides facility of customized representation of the output to the users of the system. The interface layer also accommodates other information such as local databases, log of the system and frequently asked queries. Fig. 1 illustrates these layers in a generic framework.

Using the framework one can generate artificial neural network based systems, fuzzy logic based systems and neuro-fuzzy systems at this stage. The framework is flexible enough to add reusable component of other paradigms such as Genetic Algorithms (GA). In this case, GA based systems can be developed or neuro-genetic systems can be developed using the framework.

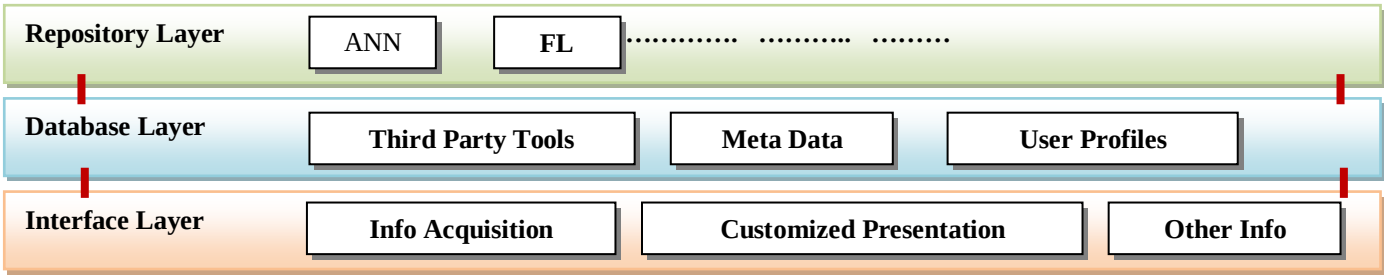


Figure 1. Multi layer framework.

A. Generation of ANN

The artificial neural network repository include code segments for feed forward, radial basis function, Kohonen self organized maps, learning vector quantization, recurrent networks, etc. Some of the codes follow existing methodology and some use novel mechanisms such as modified sigmoid and tangent activation functions. When parameters of the ANN such as number of layers and neurons, learning rate, learning algorithm, etc; the ANN is generated. To generate any feed forward, fully connected, back propagation type of multilayer ANN with supervised learning following code snippet is used.

ISBN: 978-0-9957075-2-8

```
public class NeuralNetwork
    {  protected Layer[ ] layers;
       protected int ni;
       protected LearningAlgorithm la;
       public int N_Inputs
       { get { return ni; }         }
      public int N_Outputs
       { { return layers[N_Layers - 1].N_Neurons; }
       }
public int N_Layers
       { get { return layers.Length; } }
public LearningAlgorithm LearningAlg
       { { return la; }
          set { la = (value != null) ? value : la; } }
public Layer this[int n]
       { { return layers[n]; }}
public NeuralNetwork(int inputs, int[] layers_desc,
ActivationFunction n_act, LearningAlgorithm learn)
       {     if (layers_desc.Length < 1)
      throw new Exception("PERCEPTRON : cannot build
perceptron, it must have at least 1 layer of neurone");
       if (inputs < 1)
             throw new Exception("PERCEPTRON : cannot
build perceptron, it must have at least 1 input");
                  la = learn;
                  ni = inputs;
          layers = new Layer[layers_desc.Length];
          layers[0] = new Layer(layers_desc[0], ni);
          for (int i = 1; i < layers_desc.Length; i++)
          layers[i] = new Layer(layers_desc[i], layers_desc[i
- 1], n_act);
       }



public void randomizeWeight()
       { foreach (Layer l in layers)
             l.randomizeWeight(); }
       public void randomizeThreshold()
       { foreach (Layer l in layers)
             l.randomizeThreshold();  }
       public void randomizeAll()
       { foreach (Layer l in layers)
             l.randomizeAll();  }
       public void
setActivationFunction(ActivationFunction f)
       { foreach (Layer l in layers)
             l.setActivationFunction(f);  }
       { foreach (Layer l in layers)
             l.setRandomizationInterval(min, max); }
       public float[] Output(float[] input)
       { (input.Length != ni)
throw new Exception("PERCEPTRON : Wrong input
vector size, unable to compute output value");
        float[] result;
          result = layers[0].Output(input);
          for (int i = 1; i < N_Layers; i++)
             result = layers[i].Output(result);
          return result;
       }
 }  }
```

However, users are not aware of the background code; they can interact through the framework by a friendly interface as shown in Fig. 2.

B. Generation of Fuzzy Logic

To generate fuzzy logic based system the components such as fuzzification method, defuzzification method, type reduction codes (for type 2 FL systems), etc are developed and kept in the repository layer. Many of these methods are innovative. Fig. 3 provides an overview of the components of the FL repositories.

Codes of each component shown in Fig. 3 are developed and kept ready. To generate fuzzy logic based system, an interface is generated as shown in Fig. 4.

ISBN: 978-0-9957075-2-8

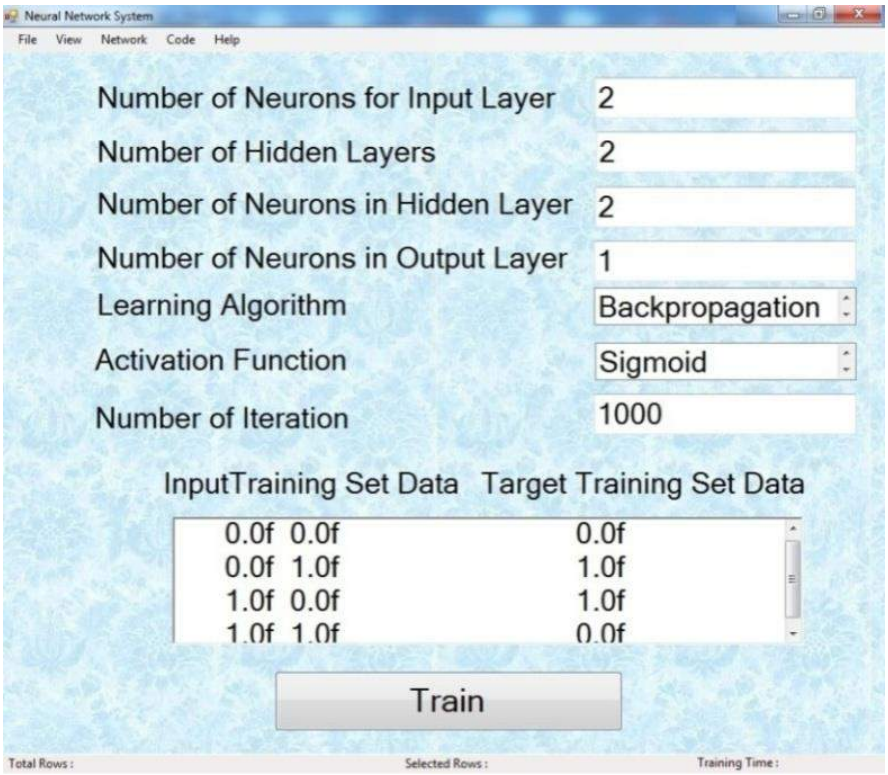


Figure 2. Interface to create ANN.

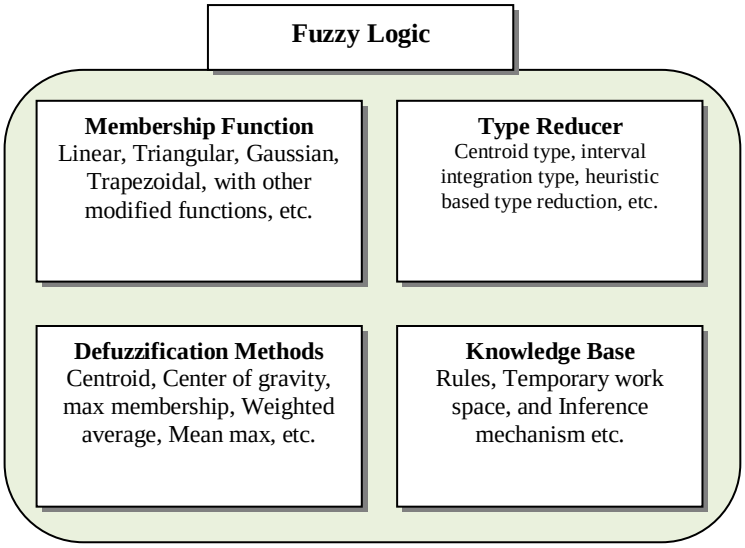


Figure 3. FL repository components.

ISBN: 978-0-9957075-2-8

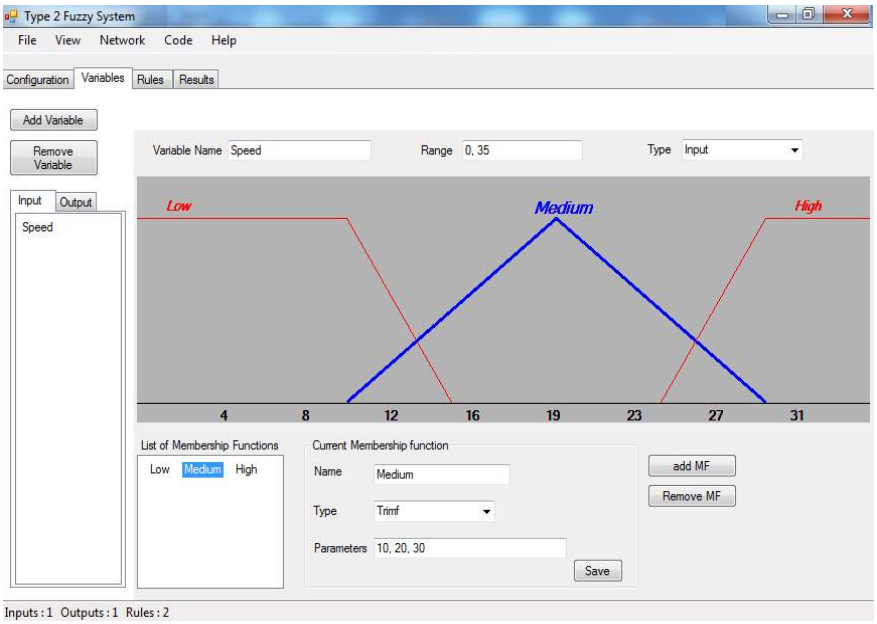


Figure 4. Generation of FL based system.

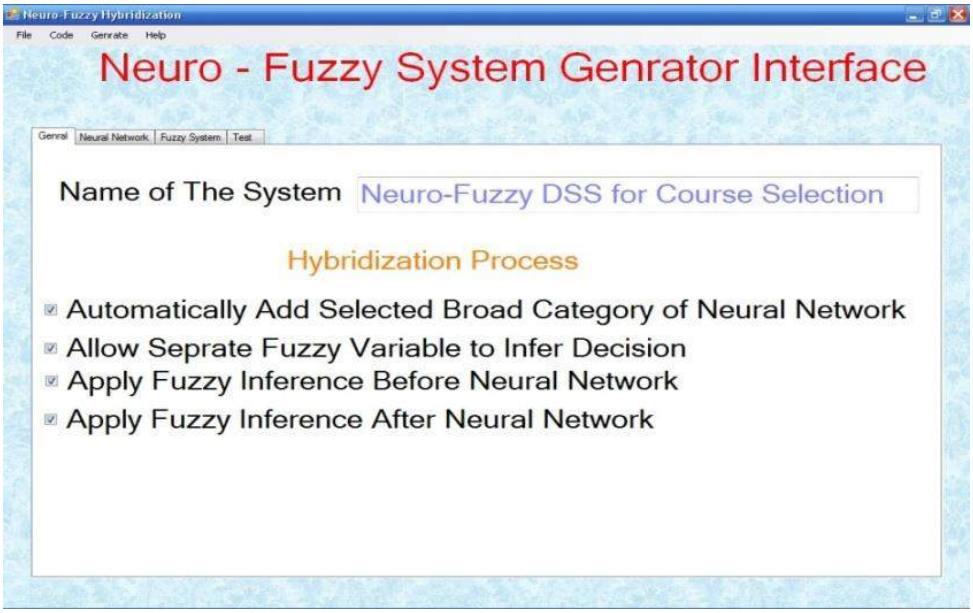


Figure 5. Developing a hybrid neuro-fuzzy systems.

ISBN: 978-0-9957075-2-8

C. Generation of Neuro Fuzzy Component

Since the components of the ANN and FL are ready they can be reused to develop hybrid neuro-fuzzy systems in different way. An interface is developed to select an appropriate type of the hybridization as shown in Fig. 5.

IV. CONCLUSION

The generic neuro-fuzzy framework presented here facilitates automatic development of intelligent system in friendly way. Using the framework many successful systems can be developed. To name a few: course selection and career advice system, student's aptitude evaluation system, neuro-fuzzy recruitment system, and portfolio management system.

It provides advantages such as reusability, modularity, and flexibility. The system developed using this framework can be saved for future use and served as good knowledge management tool. As user need not have to write codes for the components (just on demand attachment of component can be done), user may concentrate on analysis and design of the system in order to increase quality of the system. Obviously, the framework reduces the total man power required to build intelligent system, especially minimizes the role and need of computer professionals. It will also save time and save manual programming work. This generic library and framework are developed using Microsoft's Dot Net technology (Visual Studio 2010) and this can be updated easily to future release of versions. In future more components /codes can be added in to the framework along with visual interface.

REFERENCES

[1] Zadeh, A.L. (1965). Fuzzy sets. Journal of Information and Control, 8(3), 338-353.

[2] Mamdani, E.H. (1993). Twenty years of fuzzy control: Experiences gained and lessons learnt. In Proceedings of Second IEEE International Conference on Fuzzy Systems, San Francisco, CA, 339-344.

[3] Takagi, T., & Sugeno, M. (1985). Fuzzy identification on systems and its applications to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics, 15(1), 116-132.

[4] Lin, C,T., & Lee, C.S.G. (1991). Neural network based fuzzy logic control and decision system. IEEE Transactions on Computer, 40(12), 1320-1336.

[5] Jang, R. (1992). Neuro-fuzzy modeling: Architectures, analyses and applications. PhD Thesis, University of California, Berkeley.

[6] Bherenji, H.R., & Khedkar, P. (1992). Learning and tuning fuzzy logic controllers through reinforcements. IEEE Transactions on Neural Networks, 3(5), 724-740.

[7] Nauck, D., & Kruse, R. (1997). Neuro-fuzzy systems for function approximation. In Proceedings of 4th International Workshop on Fuzzy-Neuro Systems,Germany.

[8] Tano, S., Oyama, T., & Arnould, T. (1996). Deep combination of fuzzy inference and neural network in fuzzy inference. Fuzzy Sets and Systems, 82(2), 151-160.

[9] Sulzberger, S.M., Gurman, N.N., Vestli, S.J. (1993). Fun: Optimization of fuzzy rule based systems using neural networks. In Proceedings of IEEE Conference on Neural Networks, San Francisco, 312-316.

[10] Kasabov, N., & Qun, S. (1999). Dynamic evolving fuzzy neural networks with 'm-out-of-n' activation nodes for on-line adaptive systems. Technical Report TR99/04, Department of Information Science, University of Otago.

[11] Feng, J.C., Teng, L.C. (1998). An online self constructing neural fuzzy inference network and its applications. IEEE Transactions on Fuzzy Systems, 6(1), 12-32.

[12] Cihan H. Dagli, Victor Pulla, Xavier Serrano, Modeling of a Neuro Fuzzy System to Develop an Efficient Method to Get a Specific Color Paint from the Color Model Cyan, Magenta and Yellow (CMY) under Terms of Open Source, Procedia Computer Science, Volume 61, 2015, Pages 486-491, ISSN 1877-0509 , Complex Adaptive Systems San Jose, CA November 2-4, 2015, http://dx.doi.org/10.1016/j.procs.2015.09.196.

[13] Abraham, A. (2005). Adaptation of fuzzy inference system using neural learning. StudFuzz 181, 53-83, http://www.softcomputing.net/nf_chapter.pdf

[14] Jang, S.R., Sun, C.T., & Mizutani, E. (1997). Neuro fuzzy and soft computing. Prentice Hall of India Ltd, India.

[15] Mendel, J.M. (2007). Type-2 fuzzy sets and systems: An Overview. IEEE Computational Intelligence Magazine, 2(2), 20-29.

[16] Wu, H. & Mendal, J.M. (2002). Uncertainty bounds and their use in the design of interval type-2 fuzzy logic system. IEEE Transactions on Fuzzy Systems, 10(5), 622-639.

[17] Olivas, E.S., Guerrero, J.D., Valls, G.C., López, A.J., Maravilla, J.C. & Chova, L.G. (2003). A low-complexity fuzzy activation function for artificial neural networks. IEEE Transactions on Neural Networks, 14(6), 1576-1579.

[18] Su, C.L., Chen, C.J., & Yang, S.M. (2010). A self-organized neuro-fuzzy system for stock market dynamics modeling and forecasting. WSEAS Transactions on Information Science and Applications, 7(9), 1137-1149.

[19] John, R.I., & Coupland, S. (2007). Type-2 fuzzy logic: A historical view. IEEE Computational Intelligence Magazine, 2(1), 57-62.

[20] Castillo, O., & Melin, P. (2008). Type 2 fuzzy logic: Theory and applications. Springer.

[21] Malkawi, M., & Murad, O. (2013). Artificial neuro fuzzy logic system for detecting human emotions. Human-Centric Computing and Information Sciences, 3(3).

[22] Nie, D., Wang, X., Shi, L., & Lu, B. (2011). EEG-based emotion recognition during watching movies. In Proceedings of International IEEE EMBS Conference on Neural Engineering, Cancun, Mexico, 667-670.

[23] Bouzaidaa, S., Saklya, A., & Sahlia, F.M. (2014). Extracting TSK-type neuro-fuzzy model using the hunting search algorithm. International Journal of General Systems, 43(1), 32-43.

[24] Azriyenni, & Mustafa, M.W. (2013). Performance neuro-fuzzy for power system fault location. International Journal of Engineering and Technology, 3(4), 497-501.

AUTHOR PROFILE

Priti Srinivas Sajja is a Professor at P. G. Department of Computer Science,Sardar Patel University, India since 1994. She specializes in Artificial Intelligence, soft computing and multiagent systems. She is co-author of Intelligent Techniques for Data Science (2016); Intelligent Technologies for Web Applications (2012) and Knowledge-Based Systems (2009) published at Switzerland and USA apart from four books published in India. She is supervising work of a few doctoral research scholars while six candidates have completed their Ph.D. research under her guidance. She was Principal Investigator of a major research project funded by UGC, India. She has 171 publications in books, book chapters, journals, and in the proceedings of national and international conferences out of which five publications have won best research paper awards.

An Elite Model for COTS Component Selection Process

Asif Irshad Khan
Department of Computer Science, FCIT
King Abdulaziz University
Jeddah, Saudi Arabia

***Abstract*—Component-based software development (CBD) promises development of high-quality trustworthy software systems within specified budget and deadline. The selection of the most appropriate component based on specific requirement plays a vital role for high-quality software product. Multi-Agent software (MAS) engineering approach played a crucial role for selection of the most appropriate component based on a specific requirement in a distributed environment. In this paper, multi agent technique is used for component selection. A semi-automated solution to COTS component selection is proposed. It is evident from the result that (MAS) plays an essential role and is suitable for component selection in a distributed environment keeping in view of the system design and testing strategies.**

Keywords- Component selection, Component based development, COTS, Multi Agent Software Engineering

I. INTRODUCTION

Developing high-quality software product within budget and time is always a challenging concern to the software industries. Component-based software development plays a crucial role in addressing challenges faced by the software industries. CBD supports development of software using software components, also, known as commercial of the shelf system (COTS). COTS components are sold in open market and manufacturers of COTS define how a component can plug and play into a system based on the specified requirement [1].

Ideally, COTS components are like black box solutions, one has to learn how a component can be plugged into the system by using its required and provide interfaces as defined by its manufacturers. The component is developed by following software development phases such as System requirement specification, System Modeling and Design, System implementation and Testing [10]. COTS Components are highly reliable and trustworthy as they are rigorously tested by its manufacturer and most of the components have a rating that is evaluated by its customers [12].

However, Selection of component plays a vital role in developing high-quality software systems within specified constraints like budget, efforts and development time. Selection of candidate component as per specified requirement is a challenging task as the selection is usually based on multiple criteria such as functionality support, cost, reliability, security, adoptability etc.

Multi-Agent System (MAS) is intelligent software system consists of software components also known as Agent. These agents can interact with each other or can coordinate with each other to accomplish assigned task. For example, agents can be used to identify most suitable candidates for a particular set of requirements.

Multi-Agent approach is ideal for this research as the agent can explore all available servers in the distributed network and recommend potential candidate components.

This paper is organized as follows: section I describes introduction of the paper, sections II discusses about the related work, section III puts forward the proposed model, section IV explains the role of different agents in the proposed mode, and section V conclusion and future work.

II. RELATED WORK

Software engineering research community has suggested different strategies and methodologies for the selection of candidate COTS components.

Rikard et al. surveyed component selection state of the art in the embedded systems domain and presented a meta-model for selection method based on common activities and practices. The authors also gave suggestions which can be utilized as a schema when making the technique and methodology for component selection [2].

ISBN: 978-0-9957075-2-8

X. Burgu´es et al. suggested a process model for combined selection of components based on the features of two levels, a global one matching to the combined selection itself, and a local level where all the individual selection processes take place in specific areas. The main contribution of this study is to consider the aggregate expense for a system rather than specifying in advance the individual expenses for various components [3].

Lawrence et al. proposed an approach (CARE/SA) which supports the iterative selection (matching and ranking) of COTS components. The author considered COTS component's representation within architecture as an aggregate of their functional and non-functional requirements with its own set of attributes. The proposed framework could be viewed as an extension to previous methodologies with a systematic approach to matching, ranking, and selection COTS components as shown in figure 1. [4]

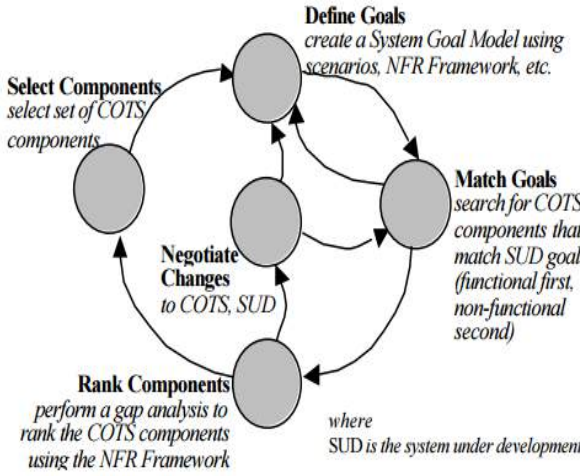


Figure 1. Overview of the CARE Process [4].

Zahid et al. in their work evaluated the recent practice of evaluation and selection of COTS components from software houses in Pakistan. Closed questionnaire methodology was adopted in the study. The result shows that no criterion is utilized for evaluating and selecting COTS components by the majority of respondents. Authors proposed a framework for COTS selection, integration and evaluation [5].

Tarawneh et al. [6] suggested a framework support and improve the COTS software evaluation and selection processes in industry. To achieve this objective the authors have shown that specific objectives have to be addressed:

a) Identify the processes which support COTS software evaluation and selection.
b) Determine the criteria or requirements which are important for successful evaluation and selection process.
c) Propose methods and techniques to address the mismatch between COTS features and customer requirements.
d) Develop a repository to manage information from previous selection cases that support the decision-making process.

M. Shakeel et al. [7] proposed component selection methodology for component-based software engineering. The author claimed that most of the IT developers in development countries (UDCs) are unaware about the COTS evaluation and selection methodology. This methodology will reduce development times, cost effective and less strenuous efforts on the basis of software quality model ISO/IEC 25010 as shown in figure 2.

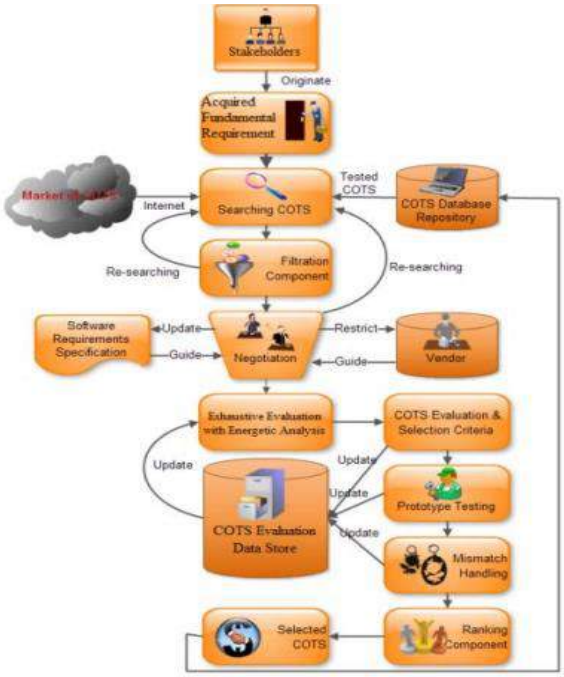


Figure 2. IROTS Process [7].

Shah et al. proposed a method for component selection using fuzzy logic. Several software quality factors like reusability, portability, security, maintainability etc were considered in the proposed method. This methodology was evaluated using hypothetical case study [8].

Agora [9] is a web-based component search prototype proposed by Carnegie Mellon software engineering Inst. Agora provides agents that crawl the

Web for components. In this model, the whole process is divided into two main tasks.

One of the task which is an automated background task is performed by an agent, i.e. a background agent is responsible to automatically search the location and indexes of COTS on the web for component specific/specification model. While the component selection and retrieval task is a manual and done by the expert as shown in figure 3.

The main advantage of agora approach is its ability to automatically build an index of available worldwide components on the web [9].

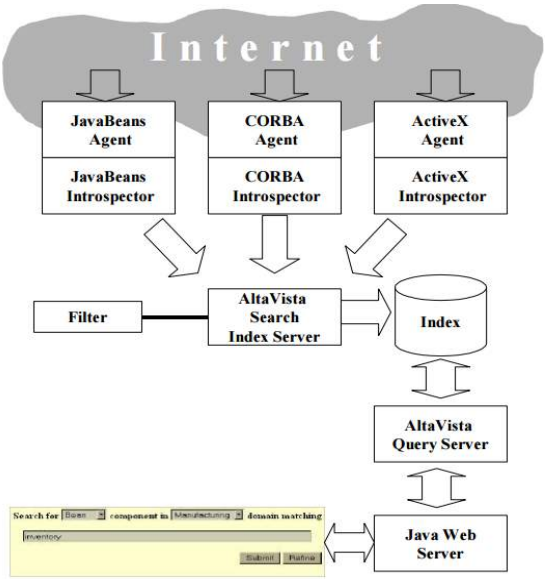


Figure 3: Agora Architecture [9].

III. PROPOSED MODEL

The currently available models for COTS component selection are complex and difficult to implement and also its time consuming. The analysis indicated that there is a need of improvement to COTS component selection process.

This paper proposed an efficient semi-automated COTS component selection technique which is easy to understand and easy to implement as well.

A semi-automated model [ABCS] for COTS component selection is proposed. This model is based on the multi-agent technique in which a group of agents works together with the single objective of completing a specific task i.e. identifying the list of suitable candidate components. Figure 4 shows a conceptual view of this approach.

The whole process of identifying the candidate component requires multiple agents to work together. A multi-agent system is a loosely coupled network of several agents which interact among themselves or with external environments to solve problems that are beyond one's individual abilities or knowledge of each problem solver.

As multi-agent is composed of several autonomous agents, there is a need of coordination and cooperation among the agents as shown in the Figure4.

Firstly the task of searching COTS component as per defined criteria is broken into several sub-tasks and each of the sub-tasks is assigned to specific agents. For example, sub-tasks like matching requirement against integration complexity, security features verification, cost-benefit analysis, adaptation, and quality assurance verification is assigned to different agents as shown in figure 4.

These agents, after solving their respective sub-tasks, co-ordinate with the main agent to produce the result's set of candidate components as per the selection criteria in the form of a weighted matrix [11]. Finally, an expert team will decide the candidacy of the most suitable component available from the list of choices.

ABCS model take the outline of the user requirement [User requirements may be in plain English or in script form or any other defined form] and application domain [like health, agriculture, business, education etc] as the input parameter and select the candidate component from the list of available components from different repositories in a distributed network as shown in figure 5.

As it is often seen that based on the requirements, there might be several COTS products that match the requirements for different degrees, but in few cases most likely none of the several candidates would completely match the user requirements.

The proposed model also has an agent integrating tool [ABIT] that takes system architecture and list of candidate components and integrates them to generate source code for integrated components as shown in figure 6.

In case of any compatibility issues it is handled by compatibility test agent. It tests the compatibility issues while integrating the selected candidate components with the system architecture.

This agent automatically verify and validate the compatibility of selected COTS component for any risk of failure since selected component has to integrate into the system and there is always a risk of component failure at later stages of development.

In case the test shows compatibility problem while integrating the selected candidate components, the system will select the next possible candidate component from the ABCS list as shown in figure 6. Interactions between these agents are shown in figure 7.

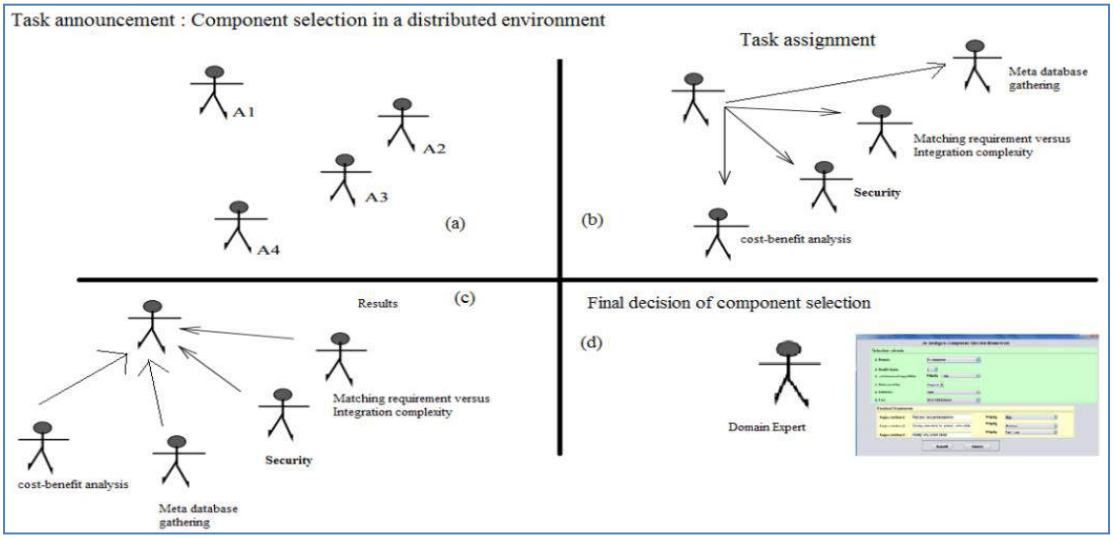


Figure 4. Conceptual view of multi-agent approach for component section.

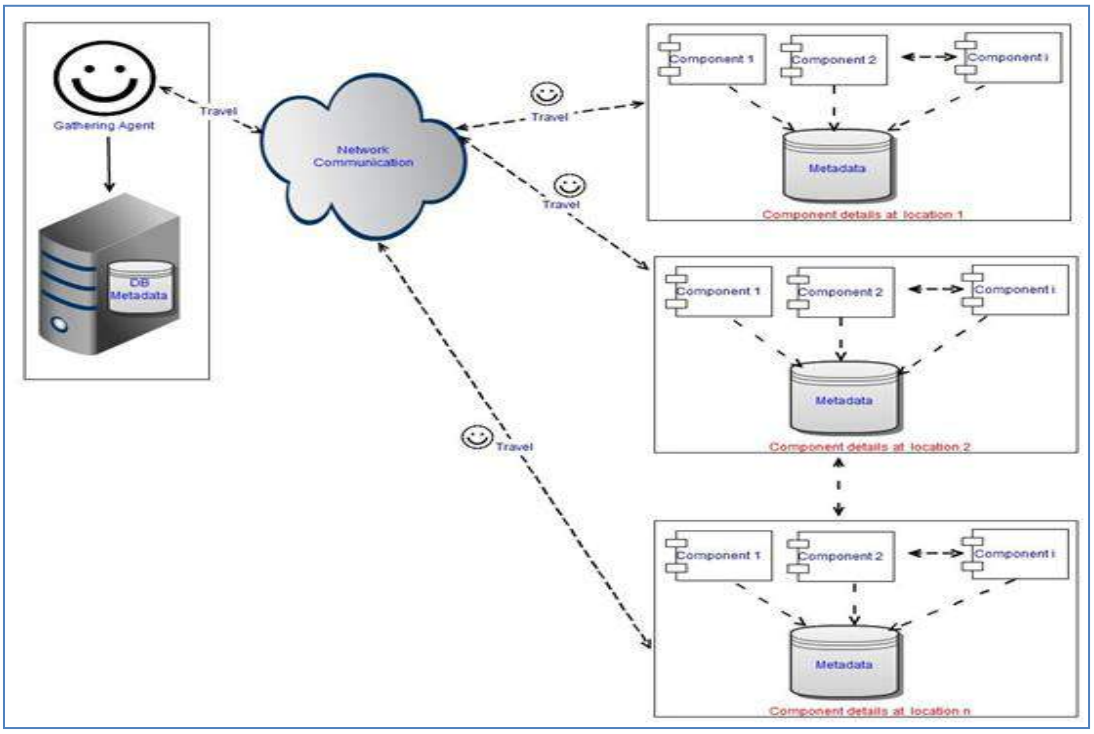


Figure 5. Conceptual Model of Component Gathering Agent.

IV. Agent Role In The Model

A. Gathering Agent

The main responsibility of this agent is to maintain in-house component meta-data database. Once requirement is specified to this agent, it travels in distributed network, collect meta-data details related to available matching components and classify them based on the component model to update in-house database.

Vendor of the component maintain meta-data details in online repositories, as shown in figure 5. Following are some of the listed categories of component classified based on the component model.

- Microsoft's COM/DCOM/COM+ components.

ISBN: 978-0-9957075-2-8

- Microsoft's .Net components.
- Object Management Group OMG's CORBA.
- Oracle J2EE or JEE 5 components
- Oracle JAVA RMI components

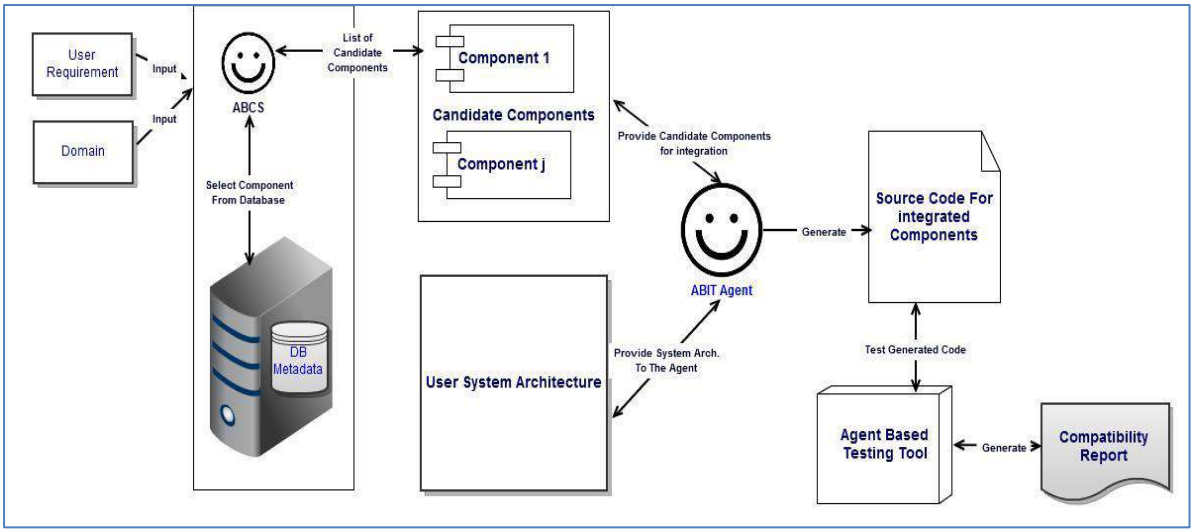


Figure 6. Model of Multi-agent based Component selection and Integration with Testing Tool.

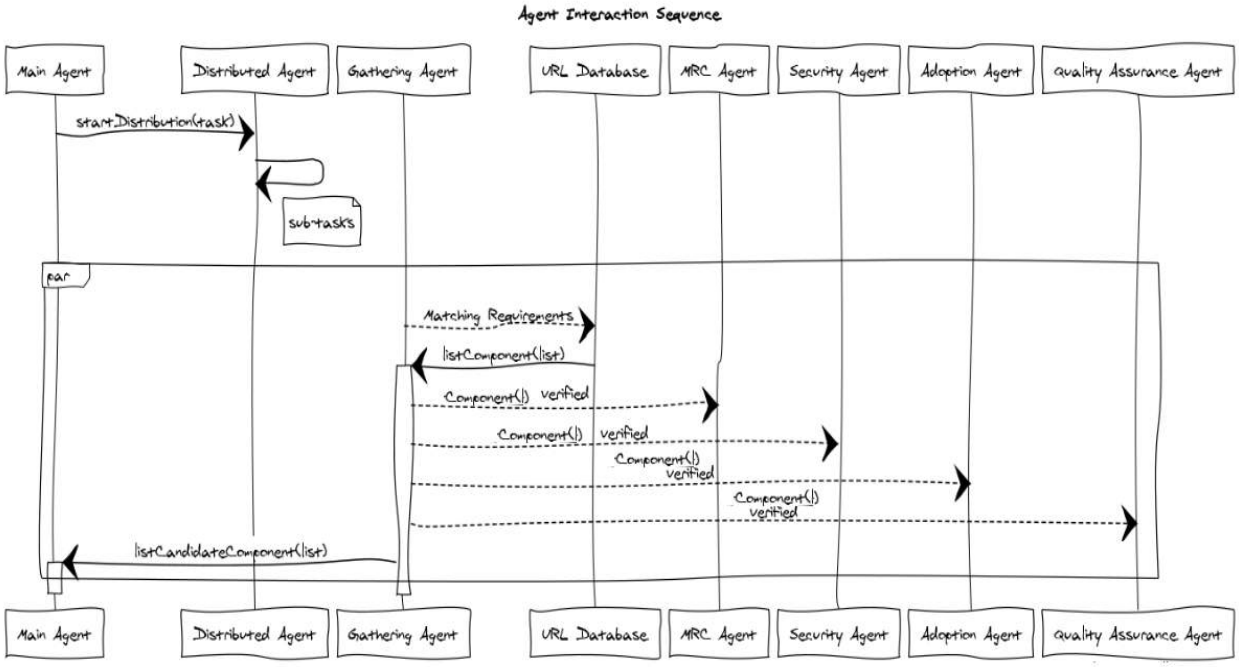


Figure 7. Agent Interaction Sequence diagram for Multi-agent based Component selection.

B. *MRC Agent*

Matching requirement versus Integration complexity is addressed by MRC agent. It plays a major role in the evaluation of every component based on its matching

requirement and its relative complexity to integrate.

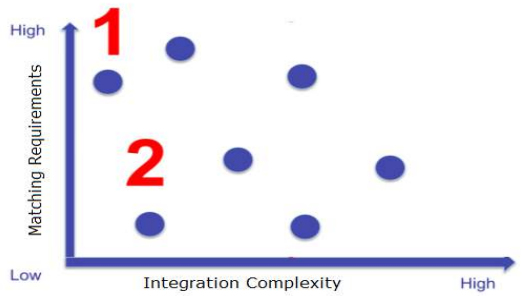


Figure 8. MRC Agent.

The component that has highest matching requirement and lowest effort to integrate will be considered from the component candidate list as shown in figure 8 and figure 9.

Recommended Candidate Components

Component name	Matching index	Recommendation
Component A	0.76	
Component B	1	High
Component C	0.82	
Component D	0.76	
Component E	0.95	High

Figure 9. MRC Agent.

C. *Security Agent*

The role of this agent is to verify the security features in the selected candidate components. The components should be verified for any breach of security. This agent has authentication and authorization manager which checks how certificate authority, usable security policies etc. are defined and security mechanisms are implemented and supported.

D. *Cost Agent*

The main responsibility of this agent is to take part in the cost-benefit analysis so as to compare the component features & functionalities with the cost. Costs should be seen in broad aspects; because low cost components may result in higher total cost along the product's life cycle. Buying component with loads of features also results in a higher cost and most of the features remain unused by the users.

E. *Adoption Agent*

The main responsibility of this agent is to verify component adoption process by studying all pre-requisite in a way that components can be easily integrated into the system architecture effectively.

F. *Quality Assurance Agent*

It is well-known fact that no software can be 100% error-free. To check component high-quality features, knowing that it is developed by the vendor company and may have some issues, it is necessary to check the rating of the component as most of the vendors maintain a rating of the components.

Also, it is necessary to find out what problems other users are encountering from vendor site where issue tracker related to the component are posted. The role of this agent is to validate and verify above mentioned high-quality attributes and generate a software quality metric for measurement that is used to evaluate software quality in a system as shown in figure 10.

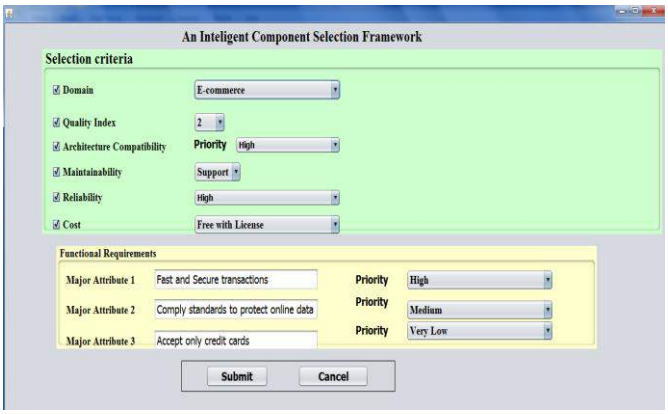


Figure10. MRC Agent.

V. CONCLUSION

In this paper, an approach to COTS component selection was presented. The software component selection is done based on different attributes like availability of features and functionalities in the component as per requirements, component adoption easiness in the architecture, security, cost, quality of service etc.

The proposed approach attempts to find the best candidate component list based on the requirements in a distributed environment. The multi-agent technique is considered in this approach by breaking a task into several sub-tasks and these sub-tasks are assigned to specific agents to be solved.

All the rigid conditions are taken into consideration and based on this; the most optimal component sets are generated. Finally, domain experts will decide the candidacy of the most suitable component available from the list of choices.

This approach gives acceptable results. This semi-automated solution to COTS component selection has the main improvement in reduction time of COST selection process which leads to acceleration of development and time to market. Further, knowledge of available COTS component is another advantage of the proposed solution.

ISBN: 978-0-9957075-2-8

REFERENCES

[1] M. Elammari and R. Ali, "Towards a Multi-Agent Mechanism for Software Component Selection", The International Arab Conference On Information Technology (ACIT2014), Zarqa University in Jordan.

[2] Rikard Land, Laurens Blankers,Michel Chaudron, Ivica Crnković, "COTS Selection Best Practices in Literature and in Industry", Mälardalen University, School of Innovation, Design and Engineering, Västerås, Sweden.

[3] X. Burgués, C. Estay, X. Franch, J.A. Pastor, and C. Quer, "Combined selection of COTS components", Proceedings of ICCBSS, February, Orlando, Florida USA, 2002, pp. 54–64.

[4] Lawrence Chung, Kendra Cooper, "COTS-Aware Requirements Engineering and Software Architecting", Department of Computer Science, University of Texas at Dallas.

[5] Zahid Javed, Ahsan Raza Sattar, Salman Afsar, Muhammad Shakeel Faridi, "An Empirical Study of COTS components Persuasion, Evaluation & Selection and Integration in software houses Faisalabad, Pakistan", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, November 2012.

[6] Feras Tarawneh, Fauziah Baharom, Jamaiah Hj. Yahaya and Faudziah Ahmad, "Evaluation and Selection COTS Software Process: The State of the Art", International Journal on New Computer Architectures and Their Applications, The Society of Digital Information and Wireless Communications, (IJNCAA) 1(2): 344-357, 2011 (ISSN: 2220-9085)

[7] M. Shakeel Faridi & Zahid Javed, M. Haris Abid, Mudassar Ahmed, Dr. Md Asri Bin Ngadi, "IROTS: A Proposed COTS Evaluation & Selection Methodology for Component Based Software Engineering in Under-Development Countries", Department of Komputing Universiti Teknologi 2nd International Conference on Advances in Computer Science and Engineering (CSE 2013), Malaysia, Johor Bahru, Malaysia

[8] S Nazir, MA Khan, S Anwar, H Khan, M Nazir "A novel fuzzy logic based software component selection modeling", 2012, International Conference on Information Science and Applications, 1-6.

[9] R.C. Seacord, S.A. Hissam, K.C. Wallnau, "Agora: A Search Engine for Software Components", IEEE Internet Computing, vol. 6, no. 2, pp. 62-70, 1998.

[10] Khan, A.I., Khan, U.A., 2012. "An Improved Model for Component Based Software Development." Software Engineering 2, no. 4 (2012): 138-146.

[11] Khan, A.I., Alam, M.M., Shariq, M., May 2015. " A Perspective Study of Intelligent System for Component based Development". International Journal of Computer Applications 117(4):11-17.

[12] A. Barnawi, M. Rizwan Jameel Qureshi and A. IrshadKhan, "A framework for next generation mobile and wireless networks application development using hybrid component based development model", Int. J. Res. Rev. Next Gener. Netw. (IJRRNGN), vol. 1, no. 2, (2011), pp. 51-58.

AUTHOR PROFILE

Asif Irshad Khan, Ph.D., is working as a faculty member in the department of Computer Science, FCIT, King Abdulaziz University, Jeddah, Saudi Arabia. Twelve years of experience as a professional academician and researcher. Dr. Khan received Ph.D. in Computer Science and Engineering from Singhania University, Rajasthan, India and Master & Bachelor degrees in Computer Science from the Aligarh Muslim University (A.M.U), Aligarh, India. He has published several research articles in leading journals and conferences. He is a member of the editorial boards of international journals and his current research interest includes Software Engineering with a focus on Component Based and Software Product Line Engineering.

Game Development - Bounty Rescuestep

Prasad B
Dept. of CSE
MLRITM, Jawaharlal Nehru Technological University
Hyderabad, India.

M Sai Kumar Reddy
B.Tech, Dept. of CSE
MLRITM, Jawaharlal Nehru Technological University
Hyderabad, India.

V Srikanth Reddy
M.Tech, Dept. of CSE
MLRITM, Jawaharlal Nehru Technological University
Hyderabad, India.

R Raja Kishore
Dept. of ECE
MRIET, Jawaharlal Nehru Technological University
Hyderabad, India.

***Abstract* - Now a day everyone is having a passion of playing Survival games so, keeping on mind this project gives a best experience on playing the best escaping game. Usually this game consists of very high visual graphics. Ball is the main player in the game designed, the main aim of the player is to hold its life and reach from source to destination by escaping from the different disturbances like obstacle which come in between. Game has many modules mainly Ball control, Checkpoints, Enemy kill and die, Self-destruct, Coin pickup and many other modules. Game has made for all devices like, windows and Mac OS systems. This Game is made by Unity Platform, and programming is with Javascript. Entire game gives a good experience with good Visual Effects.**

Keywords: Survival games, Escaping game, Javascript, Visual Effects.

I. INTRODUCTION

Bounty Rescuestep is game which gives the more challenging levels than the other survival game in the world. The game is about the ball surviving from the different types of obstacles in the 2 levels of the game using life expectancy. Bounty Rescuestep gives a gives a best visual design, simple controls and best comprehension of the development of 3D objects in the game. It provides high difficulty in playing with the obstacles and escaping from them. It gives high trouble in playing with the obstacles and getting away from them.

Bounty Rescuestep is created by a gaming platform called "Unity 3d 5" and animation of gaming objects like Ball Destroy, Smoke effect and Ball pieces are done with the animation software called "Blender 2.0". The programming is done to the objects like Ball Movement, Ball Kill, Coin Pickup and Enemy Kill are done with the Programming language "Java Script (js)". 3D modeling, Rendering of the objects and physics applied on the objects are done with the use of Unity.

Bounty Rescuestep is made for the different platforms like Windows, Linux and Mac PCs. Controls are made only for PCs which has a keyboard because, all controls are programmed for only keys pressed for arrow keys and space. By starting the game asks frequently for the setting like Resolution, Full screen or windowed, Graphics quality. If these settings are done then game starts.

II. SURVEY

Motivation: Maximum of people are spending their time in playing games in free time and that too a survival games have got a huge popularity and I am from those people who plays a lot of survival games. Got an interest to create a game like a survival game to increase design and creativity skills in me and give a challenging game to world. Aim the addiction of people on the game.

Scope: Gaming is a passion in every person's life. Not only in designing a game, but also playing games will gain some ideas and creativity in the person. More Challenging levels and difficulty will increases the anxiety in the person's mind to play the game more.

Objective: Bounty Rescuestep is a game of huge escaping techniques and having the challenging levels in it. Its main objective is to reach the destination using the life span of a ball. Bounty Rescuestep aims the people who are addicted to the survival games and gives full satisfaction and makes people to addict to this game.

Problem Definitions: Problem in the existing system is failed to give a good visual graphics and challenging levels. The existing system Nokia Bounce game has these problems and light weighted obstacles are the main drawback of the game. These problems are analyzed and we have given a possible outcome to the problems defined.

Existing System: Nokia Bounce game is a classic game and got popular in 2009. Its Gameplay is about to survive a ball from different obstacles like water, lamp and rings. In the part of gameplay there will be power up ring to make ball bigger. Gamepley will be exceptionally appealing and the levels will be anything but difficult to play. It is a 2D game so the visual graphics will be in medium and the sprite's pixels also will be visible because of low visual graphics. Bounce game was very much popular because, it was the attractive game in those days in the Nokia mobile phone. After the completion of J2ME mobile generation the new technology Android has launched many games in 3D and also in 2D. There are many different levels to survive in the game and keep alive till it reaches the destination. There were many obstacles like spiders, sharps and fire. If the ball hits to these obstacles then the ball loose its health and then restarts the game from the checkpoint which was reached last In these last 7 years the developer of Bounce

ISBN: 978-0-9957075-2-8

game was kept silent by leaving it with all the bugs but, people can't forget the game and its popularity. Keeping that on mind this project is started with the more advanced features and high visual graphics.

Disadvantages of Existing System: Nokia Bounce game has many disadvantages as it was developed many years ago. Now new technologies have been developed to solve those disadvantages and give a adventurous game to the users. The main disadvantages of the Nokia Bounce game are:

- Bounce game has very less difficulty levels to survive and very easy to play the game and complete it.
- Bounce game has very less visual graphics and designed with old technology 2D.
- Obstacles are not exceedingly perilous to survive and we can't kill them in the gameplay.

Pixel clarity in the diversion will less and every one of the items in the game is not obviously outlined.

III. DESIGN

Systems design is the way toward characterizing the design, segments, modules, interfaces, and information for a systems design to fulfill determined prerequisites. Systems design could be viewed as the utilization of systems design hypothesis to item advancement. There is some cover with the controls of frameworks examination, design and frameworks building.

3.1 System Proposed

Bounty Rescuestep is a game which is designed with the 3D technology with high visual graphics and challenging levels. Diversion will need to trouble levels however numerous checkpoints scattered amidst the levels. Checkpoints are utilized to restart the diversion from the closest indicate abstain from beginning from the earliest starting point.

Bounty Rescuestep has designed with 3D animated enemies and Obstacles. Enemies are scatters in the game at every step to be made more challenging for player. Obstacles are outlined in a manner that, ball can't escape from those hindrances effortlessly. Many obstacles designed like Hammers, Fire Balls, Lamps, Sharps, rockets and other unusual obstacles in the game.

Unusual obstacles are like, when the ball is going on unexpectedly the way will be destroyed and ball will fell down and the game starts again. Furthermore, the ball will move and all of a sudden at a specific place a colossal divider will fall on the ball and ball will be devastated. At last, a circumstance happens where ball will be encompassed by numerous obstructions and around then an undetectable collider will detect the ball and opens an approach to escape from that circumstance. Ball Destroy effect is designed in the software called "Blender". When a ball hits to an enemy or obstacle then the ball break in to the pieces. Whenever any fire obstacles like Fire Ball or Rocket hits the ball then the ball will be blasted and breaks in to pieces and the smoke effect will be instantiated. After escaping from all these obstacles the ball reaches a destination where at end a rocket launcher will be present. The ball must destroy the Rocket launcher and reach the home to enter in to next level or to complete the game.

Advantages of Proposed System: Bounty Rescuestep has many advantages as it was developed by new platform Unity 3D 5. New technologies has been developed many advances in making a game to gain more users and reputation.

The primary favorable circumstances of the Bounty Rescuestep game are:

- Bounce Rescuestep has high difficulty levels to survive and very hard and challenging to play the game and complete it.
- Bounty Rescuestep game has very high visual graphics and designed with new technology 3D.
- Obstacles are profoundly risky to survive and we can even kill them in the gameplay however not every one of the obstacles.
- Pixel clarity in the game is great and execution and speed exceptionally productive and can change the design and determination settings before the gameplay.
- Challenging levels in the game makes very addicted to the players. Animations are with very high visuals than the existing system.

3.2 System Architecture

System architecture is the calculated model that characterizes the structure, conduct, and more perspectives of a system. A design description is a formal portrayal and representation of a system, sorted out in a way that backings thinking about the structures and practices of the system.

Entity Component System (ECS) is an architectural pattern that is mostly used in game development. An ECS follows the Composition over inheritance principle that allows greater flexibility in defining entities where every object in a game's scene is an entity (e.g. enemies, bullets, vehicles, etc.). Every Entity consists of one or more components which add additional behavior or functionality. Therefore the behavior of an entity can be changed at runtime by adding or removing components. This eliminates the ambiguity problems of deep and wide inheritance hierarchies that are difficult to understand, maintain and extend. Common ECS approaches are highly compatible and often combined with data oriented design techniques.

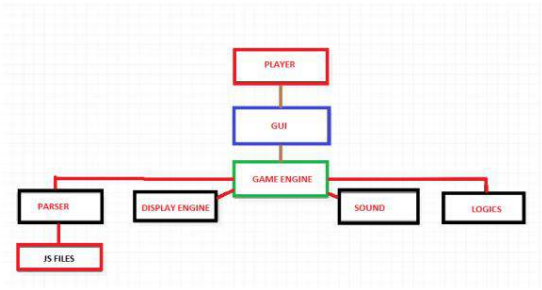


Figure 1. Architecture of the system.

ISBN: 978-0-9957075-2-8

3.3 Modules

Bounty Rescuestep is a purely dependent on the modules. All the modules are created by the language JavaScript. There are many modules in the game for every action for which the player has made.

The main modules in the game which are frequently used as follows:

- *Ball Control:* Ball control is utilized to take the input from keypad and execute as indicated by the code written in it. At the point when a player types horizontal arrow keys then the ball pushes ahead and in reverse and space is utilized to bounce. Numerous other controlling rationales have been utilized.
- *Ball Health:* Ball health is used to destroy the Ball object and instantiate the breaking effect and restart the game from check point or from the beginning of the game. It also used to produce ball blast effect.
- *Kill Script:* Kill Script is used whenever the ball hits to an obstacle then this module redirect to the Ball health and then the ball object will be destroyed. It has a function "OnTriggerEnter()" which takes any action when an object enters the collider.
- *Coin Pickup:* Coin Pickup is used whenever the ball collects the coins. The Script consists of Trigger function. When the ball touches the collider then trigger activates and the following logic will execute such that the coin gets invisible by providing coin effect and increments the score by 1.
- *Die On Hit:* Die on hit is a script used when a player kills the enemy by bouncing on the top collider of the enemy then the enemy dies with the animated object and then followed by destroy() function to destroy the Enemy center objects.
- *Check Points:* Check Point Script attached to the invisible collider when a ball reaches this check point then the module activates and store the position of x, y and z axes and start the game from the last check point.

IV. IMPLEMENTATION

Animated Enemy Flow Diagrams: In the game Enemy is animated from initial state to different states and flow of states are given as shown in fig 2.

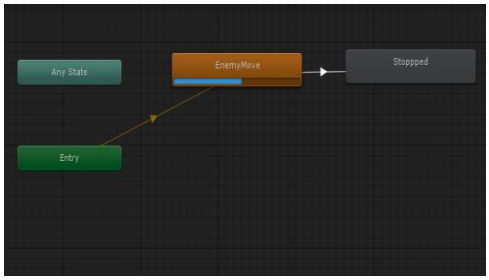


Figure 2. This flowchart represents the state when an enemy is in Moving State.

Fig 2 shows that, the blue bar represents enemy is in Moving state. Till an enemy occurs any action it loops in the same state.

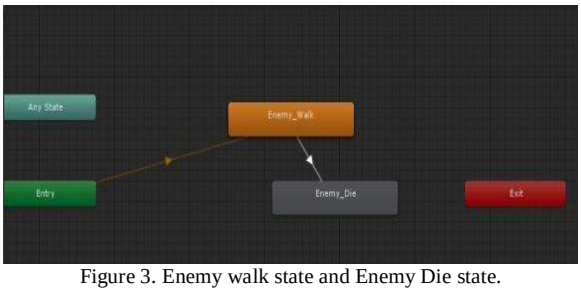


Figure 3. Enemy walk state and Enemy Die state.

Fig 3 shows 2 more states where enemy can Walk, Die and Exit. Whenever a game starts the enemy goes to "Enemy_Walk" state, if player kills the enemy then enemy goes to "Enemy_Die" State and then followed by "Exit" state to destroy the Enemy object.

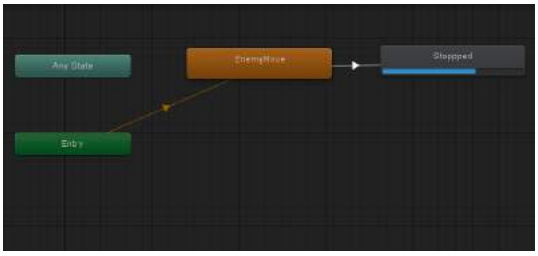


Figure 4. Enemy goes in to Stopped State.

Fig 4 shows that, the blue bar represents enemy is in Stopped state. It means that Enemy is killed and object is also destroyed.

Transformations: In the game there are many Transformations like Rotations and Positions. Objects like Coins, Enemy, Fireball and Hammers will having some actions to move horizontal and vertical. These Transformations are done as shown in below figures.

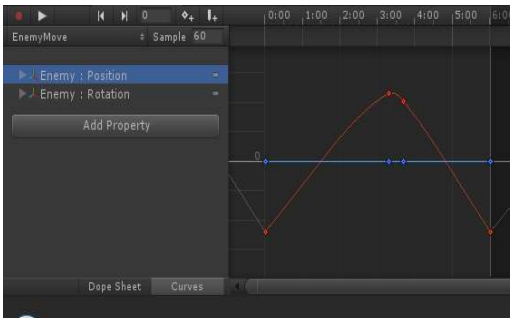


Figure 5. Change in enemy Position to move horizontal.

Fig 5 shows the position of coin within 6 seconds at 180 degrees in the Y-axis direction and add a loop to move horizontal continuously till it killed by the player.

ISBN: 978-0-9957075-2-8

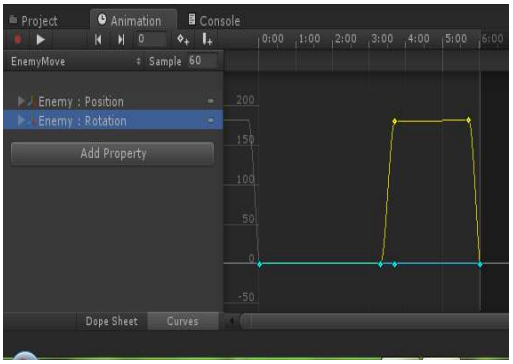


Figure 6. Graph representation of both position and rotation.

Fig 6 shows the revolution and position in a diagram representation of a foe inside 6 seconds at 180 degrees in the Y-axis bearing and add a loop to Transform. "Blue" line represents "Rotation" and "Yellow" line

V. RESULT

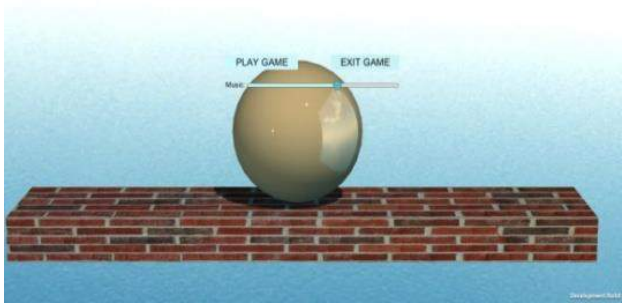


Figure 7. Main menu screen to start the game or exit.

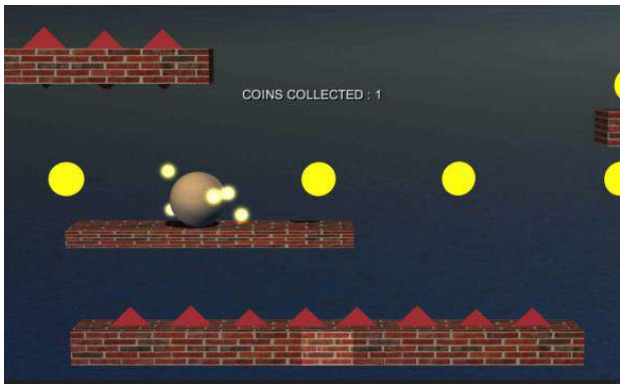


Figure 8. Ball effect when a coin is picked.

Fig 8 describes the effect which is instantiated when a ball collects the coins. The effect is called as a coin effect. Whenever the coin is picked, coin will disappear by producing this effect and the scorecard will increase by 1.

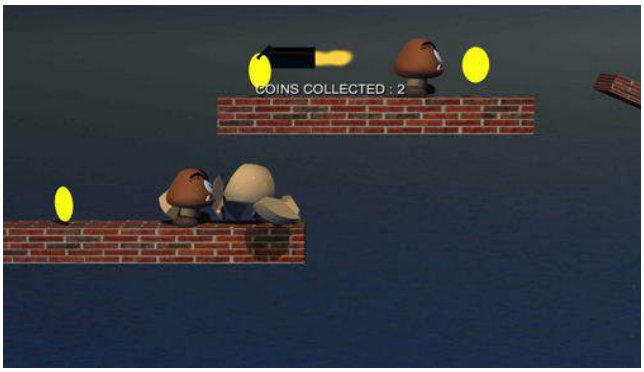


Figure 9. Ball destroy effect when a ball is hit to enemy.

Fig 9 describes, when a ball touches or hits to the enemy then ball will break in to pieces and player loses 1 life and again game will be restarted.

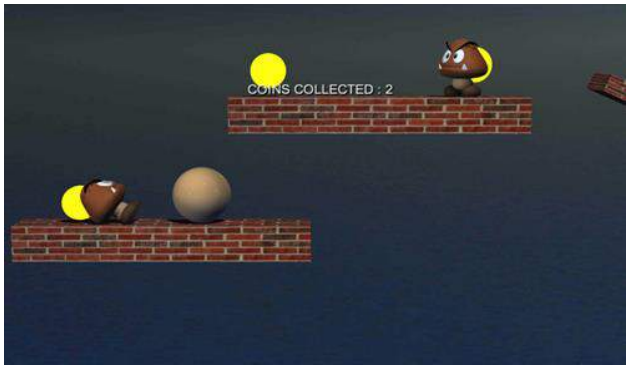


Figure 10. Effect when Ball kills an enemy.

Fig 10 describes when an enemy is killed by the ball. At whatever point a ball bounced on the highest point of the foe then an adversary will be killed as shown in the above figure.

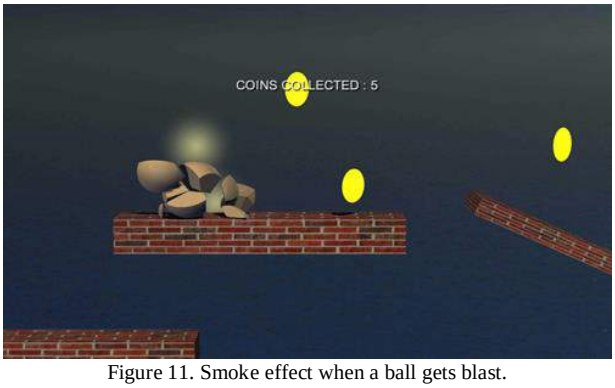


Figure 11. Smoke effect when a ball gets blast.

Fig 11 describes a blasting effect, when a rocket or any fire obstacle hits the ball then smoke effect is produced as shown.

ISBN: 978-0-9957075-2-8

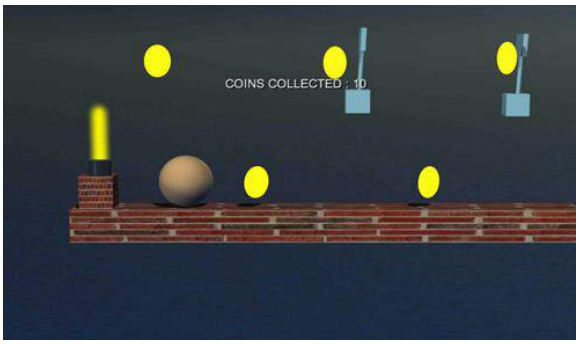


Figure 12. Light Obstacle and Hammer Objects.

Fig 12 describes 2 new obstacles called Light Obstacle and Blue color objects are called as Hammers. Even these obstacles are also can ruin the life of the player "Ball".

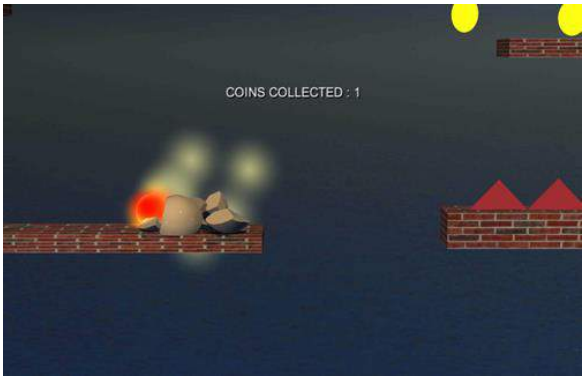


Figure 13. Ball Burning Effect when a fire ball touches the Player.

Fig 13 describes a burning effect, when a fire ball obstacle hits the ball then smoke effect is produced and breaks in to pieces as shown.

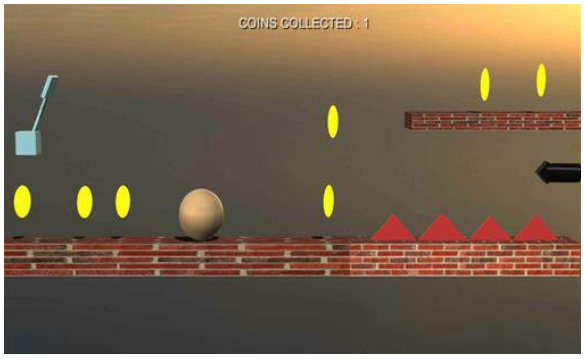


Figure 14. Clear image of Sharp of Sharp Obstacle.

Fig 14 describes a new obstacle called Sharps. Even the ball jumps or bounce on these sharps then the ball will break in to pieces and game will be restarted. The Sharps are customized to move vertically.

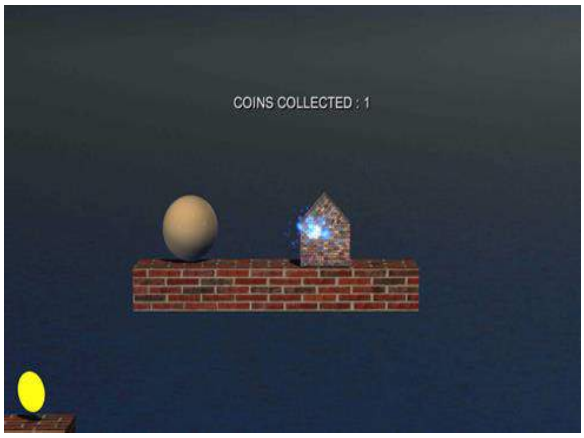


Figure 15. Final Destination called as Home.

Fig 15 describes Home, where a ball should reach final destination to start the new level or end the game.

VI. TEST CASES

User Interface: User Interface Test case is used to check that gameplay display is working under following conditions.

- Check that Full screen mode is working or not.
- Check that will the game supports all the screen resolutions are not.
- Check the graphic settings to work efficiently.
- Check for animatronics, movements of the objects, Computer graphics, all gestures like Zooming etc.
- Objects overlapping
- Character should move from specific area or screen.

Performance: Performance is measured in different conditions like the game loading speed, the total memory occupied by the game and also elapsed time of the game.

- Check the stacking time of the diversion.
- Make beyond any doubt that any move is not making impressive time; the flow in the game should be fast enough throughout the levels.

Score: Score Test case is used to check the scoring module and its functionality in the game.

- Score estimation to be measured.
- Need to monitor the levels played with the score.
- Score board to be observed.

Functionality: Functionality test case is used to check the working of logic and object in the game to work properly.

- Check diversion range , Check amusement rationale, Play till last level, Check for the reward score, Check the score climb when the level gets expanded, Menu alternatives, Different diversion modes and area, Check for the time-out.

ISBN: 978-0-9957075-2-8

VII. CONCLUSION

Our video game known as Bounty Rescuestep worked really well. It is now performing all the features that we aimed for. Now user can escape different obstacles using the player Ball. There are many levels of difficulties. We can easily change reach the destination to get through the next level.

We are getting two hundred and fifty five vibrant colours for each pixel. The limited size of the memory that is SRAM limited our design to 8 bits per pixel but still it full fills our requirement. We could get more colourful picture if we have an access to large memory.

Toward beginning of the game we planned to include more components in the diversion that is to make it for android and other cell phones. Encourage we expected to make an immaculate 3D diversion. Be that as it may, the plan of the amusement turned out to be excessively mind boggling than anticipated. It required us a great deal of investment to compose that rationale. Hence we are including these components as future upgrades.

We took in a considerable measure through this game. This project has honed our idea of VGA controller and the software-hardware interface. We took in a great deal about various memory interfaces. This project involved almost all kinds of memories that are SRAM, RAM, ROM and SDRAM. This project not only tested our technical skills but also our temperament. There were times that we practically lost trust yet we recouped through consistent focus and diligent work.

VIII. FUTURE ENHANCEMENTS

Make the game more attracted with the new difficulty levels and challenging to the user to feel more addicted to the game. Another future enhancement can be made is that game is available in android platform.

REFERENCES

[1] Walt Scacchi, "Free and open source development practices in the game community", IEEE Computer Society, No. 59, January 2014.

[2] Adams, Ernest; Rollings, Andrew (2003). Andrew Rollings and Ernest Adams on game design. New Riders Publishing. ISBN 1-59273-001-9.

[3] Bates, Bob (2004), Game Design (2nd edition). Thomson Course technology. ISBN 1-59200-493-8.

[4] Bethke, Erik (2003). Game development and production. Texas: Wordware Publishing, Inc. ISBN 1-55622-951-8.

[5] Brathwaite, Brenda; Schreiber, Ian (2009). Challenges for Game Designers. Charles River Media. ISBN 1-58450-580-X.

[6] Moore, Michael E.; Novak, Jeannie (2010). Game Industry Career Guide. Delmar: Cengage Learning. ISBN 978-1-4283-7647-2.

[7] David H. Eberly, "3D Game Engine Architecture", Elsevier, December 2004.

[8] Jason Gregory, "Game Engine Architecture", 1st Edition, 2009.

[9] Robert Nystrom, "Game Programming patterns", November 2014.

[10] Mike McShaffry, "Game Coding Complete", 4th Edition, March 2012.

[11] Joe Hocking, "Unity in Action", 1st Edition, Manning Publications, 2015.

[12] https://unity3d.com/learn - There are many ways to learn Unity. In these pages we'll find everything we need to become a Unity developer. In Tutorials we'll find video and article based content, our Documentation are a complete written manual and scripting reference, and if we'd like some time with our experts, sign up for a live Q&A session and ask your questions directly.

[13] https://www.blender.org/support/tutorials There are many ways to learn Blender. In these pages we'll find everything we need to become a Blender Animator. In tutorials we'll find video and article based content, documentation is a complete written manual and scripting reference.

AUTHOR PROFILE

Prasad B, currently working as Associate Professor in department of Computer Science and Engineering from Marri Laxman Reddy Institute of Technology & Management (MLRITM) JNTU-H, Hyderabad. Prior to coming to MLRITM worked as Assistant Professor in various universities (Lovely Professional University, JNTU-H and Pondicherry University) and have total 10.5 Years of Teaching Experience. Pursuing PhD in Content Based Image Reterival through Clustering from Gauhati University, Guwahati. Received Masters Degree (M.Tech) in Distributed Computing Systems. from Pondicherry University. Received Bachelor's Degree (B.Tech) in Computer Science and Engineering from Kakatiya University. Research areas include Data Mining, Image processing and Cryptography. Member of IAENG, IFERP, ACM, CSI, IEEE. Supervised many UG and PG projects and present guiding one DST Project. Published several papers in international and national journals and conferences. Attended various FDP, workshops. Currently working on Multi-Modal Biometric Template Security: Fingerprint and Palmprint Based Fuzzy Vault including Human Face, Eye – Iris. And also on Content-Based Image Retrieval through Clustering.

M Sai Kumar Reddy, currently studying B.Tech in department of Computer Science and Engineering from Marri Laxman Reddy Institute of Technology & Management (MLRITM) JNTU-H, Hyderabad. Member of CSI. Currently working on IoT Based Security System using Raspberry pi. Successfully Developed a 3D Game -Bounty Rescuestep.

V Srikanth Reddy, currently working as Lecture, in department of Computer Science and Engineering from Marri Laxman Reddy Institute of Technology & Management (MLRITM) JNTU-H, Hyderabad. Pursuing M.Tech in Computer Science and Engineering from JNTU-H University, Hyderabad. Received Bachelor's Degree (B.Tech) in Computer Science and Engineering from JNTU-H University, Hyderabad. Research areas include Data Mining. Member of IAENG, CSI.

R Raja Kishore, currently working as Associate Professor in department of Electronics and Communication Engineering from MLRIET JNTU-H, Hyderabad. Prior to coming to MLRIET worked as Assistant Professor in various universities in JNTU-H and have total 9 Years of Teaching Experience. Received Masters Degree (M.Tech) in Electronics and Communication Engineering from Pondicherry University. Received Bachelor's Degree (B.Tech) in Electronics and Communication Engineering from Pondicherry University. Research areas include AdHoc Networks and Image processing. Member of IAENG, IFERP, IEEE. Supervised many UG and PG projects. Published several papers in international and national journals and conferences. Attended various FDP, workshops.

ISBN: 978-0-9957075-2-8

International Journal of Engineering and Applied Computer Science (IJEACS)

A Review of Various Clustering Techniques

Ejaz Ul Haq
School of Electrical and Computer Engineering Xiamen University of Technology China.

Xu Huarong
School of Electrical and Computer Engineering Xiamen University of Technology China.

Muhammad Irfan Khattak
University of Engineering and Technology (Kohat Campus) Peshawar, Pakistan.

***Abstract*—Data mining is an integrated field, depicted technologies in combination to the areas having database, learning by machine, statistical study, and recognition in patterns of same type, information regeneration, A.I networks, knowledge-based portfolios, artificial intelligence, neural network, and data determination. In real terms, mining of data is the investigation of provisional data sets for finding hidden connections and to gather the information in peculiar form which are justifiable and understandable to the owner of gather or mined data. An unsupervised formula which differentiate data components into collections by which the components in similar group are more allied to one other and items in rest of cluster seems to be non-allied, by the criteria of measurement of equality or predictability is called process of clustering. Cluster analysis is a relegating task that is utilized to identify same group of object and it is additionally one of the most widely used method for many practical application in data mining. It is a method of grouping objects, where objects can be physical, such as a student or may be a summary such as customer comportment, handwriting. It has been proposed many clustering algorithms that it falls into the different clustering methods. The intention of this paper is to provide a relegation of some prominent clustering algorithms.**

Keywords— cluster analysis; comportment; relegation; algorithms; natural; distribution; hypothesis

I. INTRODUCTION

Data mining is an integrated field, depicted technologies in combination to the areas having database, learning by machine, statistical study, and recognition in patterns of same type, information regeneration, A.I networks, knowledge-based portfolios, artificial intelligence, neural network, and data determination. In real terms, mining of data is the investigation of provisional data sets for finding hidden connections and to gather the information in peculiar form which are justifiable and understandable to the owner of gather or mined data. The connections and hidden information gathered by data mining are represented as layouts or arrangements.

Clustering is to divide data into group of kindred objects. Objects in each cluster are homogeneous amongst themselves and dissimilar to the objects in other clusters. Fewer clusters which are representing data achieve simplification but on other side it additionally loses certain fine information and detail. It represents many fine objects by few clusters, and hence, it model data by its clusters. Cluster analysis divides data into paramount or utilizable groups (clusters). If consequential clusters are our objective, then the resulting clusters should capture the "natural" structure of the data. Cluster analysis is only a subsidiary starting point for other purposes, e.g., data compression or efficiently finding the most proximate neighbors of points. Whether for understanding or utility, cluster analysis has long been utilized in a wide variety of fields: psychology and other convivial sciences, biology, statistics, pattern apperception, information retrieval, machine learning, and data mining. In this chapter we provide a short exordium to cluster analysis. We present a brief view recent technique, which utilizes a concept-predicated approach. In this case, the approach to clustering high dimensional data must deal with the "curse of dimensionality".

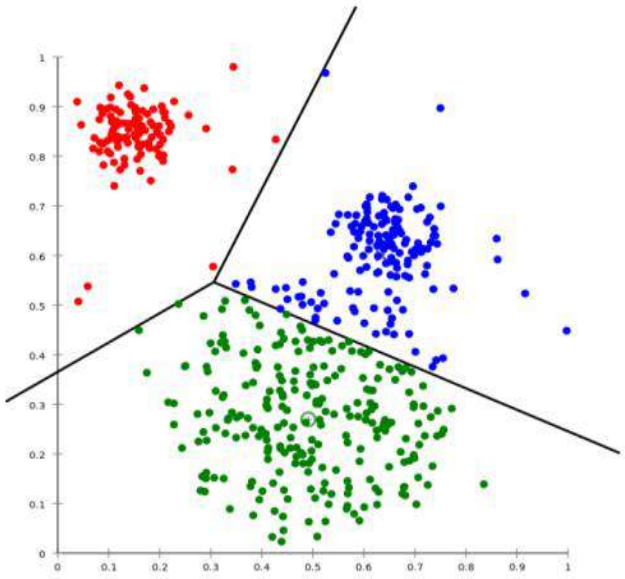


Figure 1. An illustration of making clusters.

The main achievement of clustering is allocating objects to the groups which are having similar behavior or attributes and nature, and non-likeness to rest of the instances.

ISBN: 978-0-9957075-2-8

Process of Clustering:

The overall process of cluster analysis involves four rudimentary steps as explicated below.

A. Feature Selection or Extraction

Feature selection is the process of identifying the most efficacious subset of the pristine features to utilize in clustering, whereas the feature extraction is the process of transforming one or more input features to engender incipient salient feature. Clustering process is highly dependent on this step. Infelicitous cull of features increases the involution and may result into impertinent clusters, additionally.

B. Clustering Algorithm Design or Selection

The impossibility theorem states that, "no single clustering algorithm simultaneously gratifies the three rudimentary axioms of data clustering, i.e., scale-invariance, consistency and richness". Thus it infeasible to develop a generalized framework of clustering methods for the application in the different scientific, gregarious, medical and other fields. It is consequently very consequential to cull the algorithm punctiliously by applying domain cognizance. Generally all algorithms are predicated on the different input parameters, like number of clusters, optimization/construction criterion, termination condition, proximity measure etc. This different parameters and criteria are additionally designed or culled as a prerequisite of this step.

C. Cluster Validation

As there is no macrocosmic algorithm for clustering, different clustering algorithm applied to same dataset engender different results. Even identically tantamount algorithm, with the different values of parameter engenders different clusters. Consequently it becomes compulsory to validate or evaluate the result engender by the clustering method. The evaluation criteria are categorized as:

1) Internal indices: The internal indices generally evaluate the clusters engenders by the clustering algorithm by comparing it with the data only.

2) External indices: The external indices evaluate the clustering results by utilizing the prior erudition, e.g. class labels.

3) Relative indices: As the designation suggest, this criteria compares the results against sundry other results engendered by the different algorithms.

D. Results Interpretation

The last step of clustering process deals with the representation of the clusters. The ultimate goal of clustering is to provide users with paramount insights from the pristine data, so that they can efficaciously analyze and solve the quandaries. This is still an untouched area of research.

Components of Process of Clustering:

Standard clustering methodology includes the specified Components:

(i) Pattern presentation.
(ii) Foundation of common pattern occurrence.
(iii) Collective data patterns based on likeness.
(iv) Data hiding
(v) Estimate of outcome.

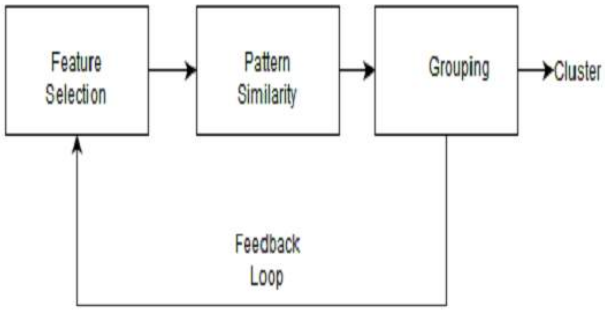


Figure 2. Component of Clustering

II. Literature Review

A. K-means versus K-means ++ Clustering Technique

This paper provides a path of computerizing k-means by selecting fluky starting midpoints with advanced efficient predictabilities. With merging k-means to a basic, flukier seeding terminology, a new articulated method that is (logk)-competitive having the optimal efficiency can be produced. This terminology guarantees an approximated ratio O (log. k) in which k is count of allied collection.

B. Consensus Clustering Based on Particle Swarm Optimization Algorithm

In presented paper, the intended accession is the PSO which used to illuminate the problem of allied collection of consensus. It is conclude the Particle Swarm Algorithm is working efficiently regarding present problem. In this paper firstly the algorithms is described which is used to create cluster as a group and consensus functions in implementation.

For building the group of clusters five distinct clustering algorithms are being used, that are K-means using the Euclidean equality schema, K-means using Manhattan equality schema, Expectation–maximization algorithm (E.M), Hierarchical schemes and P.S.O clustering. Presented algorithms generated the individual allied collections using similar data sets. By previous using the consensus method on the obtained clusters using algorithms, the labeling is done on the result of grouped clustered data.

C. Automatic Identification of Replicated Criminal Websites Using Combined Clustering

In presented paper a combined clustering method is presented which is used to link the replicated extortion websites even the criminals' use techniques to hide details. The proposed technique is used for semi-automated extortions or frauds. For this data is taken from databases of two websites that are: high yield investment programs (HYIPs) and fake-

ISBN: 978-0-9957075-2-8

escrow. After getting the data attributes of input data are extracted. Then in clustering's first stage computation of clustering is done for each input attributes by hierarchical clustering algorithm. A combined matrix is obtained on attribute basis, and then in the next stage of clustering is done with that matrix and clusters with criminal data are produced. The result implies that this technique worked efficiently as compared to general purpose methods.

D. *Fast K-Means Clustering for Very Large Datasets Based on Map Reduce Combined with a New Cutting Method*

This paper proposing a new technique in the clustering environment based on Map reducing method. A new feature is also embedding in it that is called a new cutting method. Map Reduce method helped in executing the job distributive by dividing it in to several parts and executing concurrently. By using it with K-Mean it provides facility to handle large data efficiently but the obstacle there is the increasing number of iterations which affects the overall performance. The proposed method providing solution for this obstacle by introducing a new characteristic called cutting method. By using this property, the iteration is reduced up to 30% with increasing throughput.

E. *K-Means Based Clustering Approach for Data Aggregation in Periodic Sensor Networks*

In presented paper, an optimized criteria of old PFF method is proposed named as K-PFF. In the proposed methodology mean algorithm of clustering is embedding and it is applied before the older PFF technique is applied on the generated clusters. By using K-mean the iteration of comparisons are reduced for finding the similar data. Hence it resulted in the reduced overhead of network and also reduced data latency.

F. *Extensions of Kmeans-Type algorithms: A New Clustering Framework by Integrating Intracluster Compactness and Intercluster Separation*

In presented paper, a chain of algorithms of clustering by expanding the current traditional k-means is suggested by merging the intra cluster likeness and inter cluster division. The features and effectiveness of proposed algorithms are experimented on different real-life data sets. The presented paper includes the under defined phenomenon: 1) the 3 proposed new judicious criteria's are rely upon the classical k-means, W-k means, and AWA; 2) rest resembling updated axioms are made and the concurrence is proven and 3) empirical practical's are performed to analyzed the working efficiency.

G. *Map-Reduce Processing of K-means Algorithm With FPGA-accelerated Computer Cluster*

This paper proposed an approach in which the k-means clustering algorithm is designed and implemented on an FPGA-accelerated computer cluster. The map-reduce models used with the map and reduce procedures executed paralleled by the CPU on concurrent FPGAs. In this technique two types of communication channel is used that are in first type is used for retrieval of intended instances of primary storage method which are refined through surveyors, second is the transfer of intermediate values in the mappers and reducers. By implementing k-means, system's computation and I/O functioning of FPGA era is analyzed. As compared to the hadoop environment this approach's performance is improved.

H. *Asymmetric k-Means clustering of the Asymmetric Self-Organizing Map*

In the presented paper approach of scrutiny of data is being represented which have two steps. The first step contains visualization of data which is done through asymmetric S.O.M, whereas the second step of approach is the data visualization through disorganized data that was being divided in allied collections by applying the asymmetric K-means. The outcomes of the performed work proved the effectiveness of the intended scheme upon the traditional algorithms of clustering that are the classical K-means algo, the G MM-based methodology, and DBSCAN. This approach improves the count of objects of the clusters.

I. *Data Clustering through Particle Swarm Optimization Driven Self-Organizing Maps*

In the presented paper two techniques PSO (Particle. Swarm. Optimization) and SOM (Self Organizing Maps) are combined to perform clustering task. SOM is used here for unsupervised learning which maps data patterns with high dimension into reduced mapping of low level dimensions. This reduction makes that data more efficient and better visualization is done by that tool. PSO is the intelligent technique or the optimized algorithm which work on the population which called swarm. In proposed approach, the Lbest also known as input size and Pbest are randomly chosen for each neuron particle.

J. *A fuzzy clustering algorithm to detect criminals without prior information*

The problem of recognizing criminals via communication network is resolved in this paper by proposing a technique named as a fuzzy clustering algorithm. By this algorithm, the hidden conspirators are analyzed which are not used any prior credentials. Fuzzy k means is applied on the global information. A weighted network is formed. Based on priority list, each node in the network that have link with local conjecture are mapped in to the global information cluster. This technique is applicable to large data sets as well as small data sets also. For e.g., TF-IDF method, Disease in biological network.

K. *Applying K-Means Clustering Algorithm Using Oracle Data Mining to Banking Data*

Data clustering implies the scheme of merging data into distinct collections based on the inter class features. By the collaboration data is in the structure and consequently another preparation of the data is manufactured quite simpler. The paper purposed classical k-means algorithm investigated through Data Mining with oracle. Standard scheme of

ISBN: 978-0-9957075-2-8

clustering is to apply to the eighteen attributes of 4.0 banks and 1.0 of the collective instances is produced. By obtaining the cluster, comparisons between the banks is done on the basis of defined attributes in this paper.

L. An Optimized Version of the K-Means Clustering Algorithm

The presented paper introduced an upgraded adaptation of the traditional K-Means scheme. The main focus in this paper is on the optimization of running time and that concept realized by observing the relocation of data elements that occurred at a small rate after a few iterations. So, there was no need to rejuvenate data components. The work intended here in paper establish limb on those components that are not changing their positions in relocation process and which are changing their positions.

III. Similarity Measures In Clustering

The hierarchal clustering method which is in the form of trees makes use of the equality and gap in the production of instance's clusters. For collaboration and dividing the components some specific criteria are used named as similarity. For e.g., clustering of fast food is done on the basis of calories contained, price and taste, type. Multi dimensions areas are the most significant method for evaluating the distances of objects. Researcher's main concern is with the measurement of gap rather it is obtained through the pure method or technique, or it is imitated through simulated terminology.

TABLE 1: SIMILARITY MEASURES USED IN DIFFERENT ALGORITHMS

Measures	Forms	Examples
Minkowski distance	$\left(\sum_{i=1}^{n} \lvert x_i - y_i \rvert^p\right)^{1/p}$	Fuzzy c-means
Euclidean distance	$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\lVert x_i - v_j \rVert\right)^2$	K-means algorithm
City-Block distance	$\sum_{j=1}^{k} \lvert a_j - b_j \rvert$	Fuzzy Art
Sup distance	$d_{ij} = d(\{X_i\},\{X_j\}) = \lVert X_i - X_j \rVert^2.$	Fuzzy c-mean with sup norm
Cosine Similarity	$D_C(A,B) = 1 - S_C(A,B)$	Used in Document Clustering
Mahalanobis distance	$d(\underline{x},\underline{y}) = \sqrt{\sum_{i=1}^{N} \frac{(x_i - y_i)^2}{s_i}},$	Clustering algorithms which are Hyper ellipsoidal

IV. Clustering Algorithms

The clustering algorithms are classified on the basis of clustering models. The algorithms are many in numbers but not all the algorithms are correct. The algorithms are chosen for a specific problem on the basis of experimental study. The Classification of clusters is explained in the following section.

A. Partitional Clustering

Partitional techniques engender a one-level (unnested) partitioning of the data points. If K is the desired number of clusters, then partitional approaches typically find all K clusters at once. Contrast this with traditional hierarchical schemes, which bisect a cluster to get two clusters or merge two clusters to get one. Of course, a hierarchical approach can be used to engender a flat partition of K clusters, and likewise, the reiterated application of a partitional scheme can provide a hierarchical clustering. The cluster must have two properties. They are each group must contain at least one object and each object must belongs to precisely one group.

In this type of clustering, the familiar algorithms are K-Means, K-Medoids, CLARANS, Fuzzy K-Means, and K-Modes.

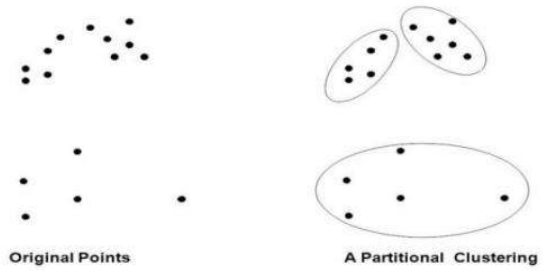


Figure 3. Partitional Clustering

B. Hierarchical Clustering

In Hierarchical type of clustering, more diminutive clusters are merged into more astronomically immense ones, or more sizably voluminous clusters are splitted into more minuscule clusters. The result of the algorithm is a tree of clusters, called dendrogram, which shows how the clusters are cognate. By cutting the dendrogram at a desired level, a clustering of the data items into disjoint groups is obtained. A hierarchy of clusters is built by hierarchical clustering. Its representation is a tree, with individual elements at one end and a single cluster with every element at the other .A hierarchical algorithm yields a dendrogram representing the nested grouping of patterns and kindred attribute levels at which groupings change.

Cutting the tree at a given height will give a clustering at a culled precision. In the above example, cutting after the second row will yield clusters {a} {b c} {d e} {f}. Cutting after the third row will yield clusters {a} {b c} {d e f}, which is a coarser clustering with fewer clusters. The merging or splitting ceases once the desired number of clusters has been composed. In general, each iteration involves merging or splitting a dyad of clusters predicated on a certain criterion, often quantifying the proximity between clusters. Hierarchical techniques suffer from the fact that interiorly taken steps (merge or split), possibly erroneous, are irreversible.

ISBN: 978-0-9957075-2-8

In Hierarchical Clustering, the familiar algorithms are AGNES, DIANA, CURE, CHAMELEON, BIRCH, and ROCK.

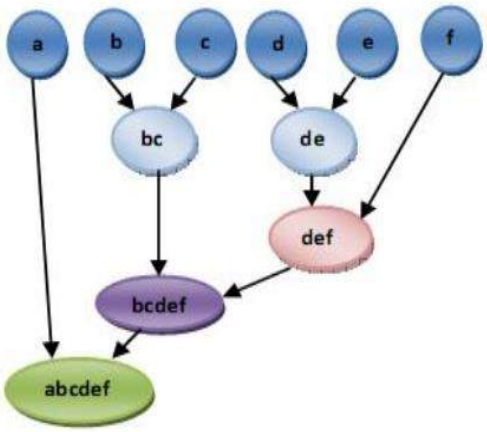


Figure 4. Hierarchical Clustering

C. Density-Based Clustering

Density-based Clusters are defined as areas of higher density than the remnant of the data set. Objects in these sparse areas that are required to separate clusters are customarily considered to be noise and border points. It requires just two parameters and is mostly in sensitive to the injuctively authorizing of the database. The quality of density-predicated clustering depends on the distance measure utilized in the function. It does not require one to designate the number of clusters in the data a priori. This method has been developed predicated on the notion of density that is the no of objects in the given cluster, in this context. The general conception is to perpetuate growing the given cluster as long as the density in the neighborhood exceeds some threshold; that is for each data point within a given cluster; the neighborhood of a given radius has to contain at least a minimum number of points. The density bases algorithms can further relegated as: density predicated on connectivity of points and predicated on density function.

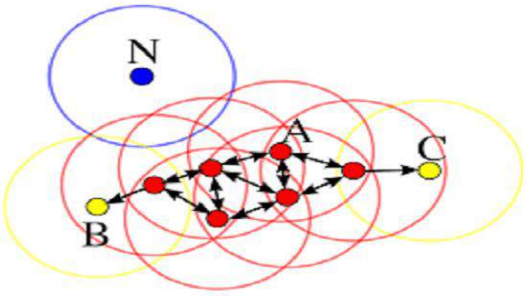


Figure 5. Density based Clustering

The algorithms in this method include DBSCAN, DENCLUE and OPTICS.

D. Grid-Based Clustering

The Grid-based type of clustering approach utilizes a multi resolution grid data structure. It quantizes the object space into a finite number of cells that form a grid structure on which all of the operations for clustering are performed. The grid-based clustering approach differs from the conventional clustering algorithms in that it is concerned not with the data points but with the value space that circumvents the data points. In general, a typical grid-predicated clustering algorithm consists of the following five rudimental steps:

- Creating the grid structure, i.e., partitioning the data space into a finite number of cells.
- Calculating the cell density for each cell.
- Sorting of the cells according to their densities.
- Identifying cluster centers.
- Traversal of neighbor cells.

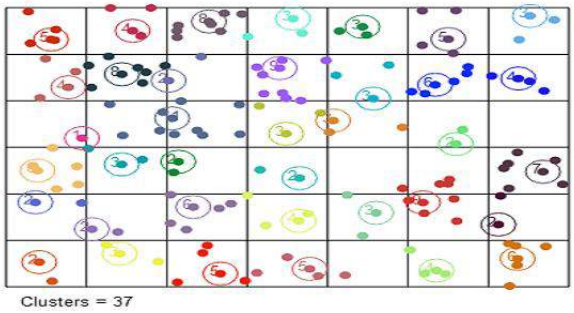


Figure 6. Grid-Based Clustering

The important algorithms in this method include STING, Wavelet and CLIQUE.

V. Problem Formulation

As we all know kmeans algorithm has some short comings which are firstly it chooses the initial seeds for center of clusters randomly which leads to wrong formation of clusters. In the presented approach a new technique is appended in kmeans algorithm to overcome these shortcomings and to reduce the iterations of algorithm. In the presented approach we implemented two new formulas by which the initial seeds for centers are selected on probability distribution basis and for calculating the distance respectively. The data points which have highest probability must be the initial center of cluster.

1. New cluster centroid using formula of average

$$V_i = \left(\frac{1}{C_i}\right) \sum_{i=1}^{c_i} x_i$$

ISBN: 978-0-9957075-2-8

2. Improved K Means ---------Distance

$$bn = \sum_{j=1}^{n} \max(d_{k-1}^{j} \quad - ||x - xj||\text{^}2,0)$$

According to these formulas, firstly we apply the cluster centroid formula to calculate the initial center of predefined clusters. Then on the basis of result of these formulas the data is distributed into clusters. Now the distance formula is applied to calculate the new distance of clusters according to new introduced formula.

These enhanced approaches provide the less iteration as compared to the classical kmeans. The error rate is also reduced to a great extent. An emerging technology which is implemented in number of fields, the basic moto of this emerging scheme is to distillate enlightenment by applying KDD to coarse data and then do the makeover into an easily accessible, ordered and understandable conformation for another use is often named as mining of data. Clustering is one of the main aspects used in mining of data. An unsupervised attainments formula which differentiate data components into count of collections by which the components in similar group are more allied to one other and items in rest of cluster seems to be non-allied, by the criteria of measurement of equality or predictability is called process of clustering. K-means is traditional clustering algorithms, but its usage with the bulk computations, make its performance quite low. The proposed schema can upgrade or boost the execution process of classical programmability of K-Means by enhancing it introducing seed selection criteria and new distance matrix method. By enhanced collaboration of these two features in algorithms, this can implement in large scale application with reduced amount of calculation and reduced iterations. The scope of the implementing terminologies in a pace originality point of view and execution span for the specific employment would be propagandize as the performance measurement criterion. This scheme's intentions are to contrap these algorithms and graphically confront the difficulties and effectiveness of the algorithm.

VI. Conclution

The main idea here is to investigate a universal efficient segregation, quick response to improved schema, of defined officials into a peculiar count of allied collections. The methodology is designed here for same kind of obstacles. With the change of segmentation obstacle like an Optimized obstacle, an improved partitioning accession is intended. After that the improved approach merged with K-means algorithm to scale the algorithm. Simulations will be performed to obtain effective execution of the improved algorithm and matched with the rest of the programs. It will help in reducing the iterations and computational time of algorithm. Also overcome the problem of increased error rate.

References

[1] Jiawei Han and M Kamber, Data Mining: Concepts and Techniques, Second Edition.

[2] Tayal Devendra K., Jain Arti, Arora Surbhi, Agarwal Surbhi, Gupta Tushar, Tyagi Nikhil (2015) "Crime detection and criminal identification in India using data mining techniques", AI & SOCIETY, 30(1), Springer-Verlag London 2014, pp. 117-127.

[3] Gonsalves Tad and Nishimoto Yasuaki (2015) "Data Clustering through Particle Swarm Optimization Driven Self-Organizing Maps", Intelligence in the Era of Big Data, Springer Berlin Heidelberg, pp. 212-219.

[4] Drew Jake, Moore Tyler (2014) "Automatic Identification of Replicated Criminal Websites Using Combined Clustering", Security and Privacy Workshops (SPW), 2014 IEEE, pp. 116-123.

[5] Agarwal Shalove, Yadav Shashank and Singh Kanchan (2012) "K-means versus K-means ++ Clustering Technique", Engineering and Systems (SCES), 2012 Students Conference, IEEE, pp. 1-6.

[6] Jassi Kaur Navjot, Wraich Singh Sandeep (2014) "An Enhanced K-Means Clustering Technique with Hopfield Artificial Neural Network based On Reactive clustering Protocol", Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference, IEEE, pp. 821-825.

[7] Rui Xu, and Donald Wunsch II , ìSurvey of Clustering Algorithms, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 16, NO. 3, MAY 2005.

[8] Pooja Batra Nagpal and Priyanka Ahlawat Mann, comparative Study of Density based Clustering Algorithms, International Journal of Computer Applications (0975 ñ 8887) Volume 27ñ No.11.

[9] Fan Changjun, Xiao Kaiming, Xiu Baoxin, Lv Guodong(2014) "A fuzzy clustering algorithm to detect criminals without prior information", Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference ,IEEE, pp. 238-243.

[10] Huang Xiaohui, Ye Yunming, and Zhang Haijun (2013) "Extensions of Kmeans-Type algorithms: A New Clustering Framework by Integrating Intracluster Compactness and Intercluster Separation", Neural Networks and Learning Systems, IEEE Transactions, 25(8), pp. 1433-1446.

ISBN: 978-0-9957075-2-8

Call for Papers

International Journal of Engineering and Applied Computer Science (IJEACS) invites authors to submit their manuscripts for categories like research papers, review articles, survey papers, technical reports , report of unsuccessful research projects , case studies , tutorials , book reviews, short communications and cross talk, extended version of conference papers for publication in our monthly issue.

IJEACS reviews manuscripts through double blind peer review process. Authors can submit their original, unpublished manuscript which is not under consideration for publication in any journal or conference through email. IJEACS publish papers in major streams of Engineering and Computer Science include following but not limited to.

Computer Science

- Software Design and Modeling
- Service Oriented Architecture
- Open Source Software
- Software Testing and Maintenance
- Software Measurement and Reliability
- Knowledge based Systems
- Image Processing and Computer Graphics
- Extreme Programming and Formal Methods
- Artificial Intelligence, Image Recognition and Bio metrics.
- Machine Learning and Computer Vision
- Algorithm Analysis and Design
- Computational Mathematics
- Data Structures and Graph Theory
- Video Coding and Data Compression
- Database Systems and Big Data Analytics
- Internet of Things, Architecture and Computing
- Parallel and Distributed Computing
- Cloud Computing, Agent Oriented System

- Communication Network and Systems
- Embedded Systems and Applications
- Cryptography and Information Assurance
- Computational Biology and Bioinformatics
- Human Computer Interaction
- Natural Language Processing

Engineering

- Micro Processor Architecture and Design
- VLSI/IC Microelectronics and Computer Design
- Parallel Processing and Digital Systems
- Wireless Transmission and Mobile Communication
- Antenna and Wave Propagation
- Semiconductors, Circuits and Signal System
- Material Science and Metallurgy
- Machine and System Design
- Robotics and Automated Control
- Manufacturing Processes and CAD/CAM
- Quality Control and Assurance
- Digital Signal Processing
- Photonics, Fiber Optics and Optical Communication
- Biosensors, Electrical-based Interactions
- Nano electronic Devices and Silicon Micro/Nano Electronics Applications
- Distributed monitoring systems and Smart Systems
- Fluid Dynamics and Implementations
- Mechanics and Vibration
- Heat Transfer
- Combustion Engines and Automobiles
- Health Instrumentations and Technologies
- Thermal Engineering
- Solar Power Systems
- Ergonomics

Submit your manuscript to: **submit@ijeacs.com** **or** **ed.manager@ijeacs.com**

Innovations continue to serve the humanity

Issue Highlights

- **Cyber-Defensive Architecture**
  Charles Kim
- **Neuro-Fuzzy Systems**
  Priti Srinivas Sajja
- **COTS Component Selection**
  Asif Irshad Khan
- **Game Development**
  Prasad B, M Sai K. Reddy, V S. Reddy, R Raja Kishore
- **Clustering Techniques**
  Ejaz Ul Haq, Xu Huarong, Muhammad Irfan Khattak

ISBN-13:978-0-9957075-2-8
00000
9 780995 707528 >

International Journal of Engineering and Applied Computer Science
Volume: 02, Issue: 01, January 2017
ISBN: 978-0-9957075-2-8