ISSN: 2454-9924

DESIGN AND IMPLEMENTATION OF SMART HOME AUTOMATION SYSTEM USING IoT

Mrs. I. Varalakshmi[1],S.Swetha[2], M.Krisha Santhoshi[3], M.Ashmitha[4], D.Souvedha[5]
[1]Assistant Professor, Department of Computer Science and Engineering
[2,3,4,5] B.Tech, Department of Computer Science and Engineering,
[1,2,3,4,5]Manakula Vinayagar Institute of Technology, Puducherry, India.
[1]varalakshmicse@mvit.edu.in, [2]swethaswami2003@gmail.com[3]santhoshikrisha22@gmail.com,
[4]ashmitha1314@gmail.com, [5]souvedhad@gmail.com

***Abstract*— Smart home system that adapts to the preferences and needs of occupants, providing a seamless and intuitive interface for controlling various aspects of the home environment. Integrate temperature monitoring capabilities to enable precise control over heating, ventilation, and air conditioning (HVAC) systems. This application is designed to enhance energy efficiency and user convenience through real-time temperature and occupancy monitoring. This system employs temperature sensorsDHT 11 for precise environmental control, activating fans to regulatethe room temperature.An infrared (IR) sensor detects human presence at the door, triggering the illumination of room lights for added security and comfort.The integration of a ESP8266 Wi-Fi module facilitates remote monitoring of temperature,humidity, IR sensor and LM 358 values. This comprehensive solution aims to create an intelligent and responsive living space, optimizing energy usage and providing users with a seamless and connected home experience.**

***Keywords* — IoT, Sensors, Automation, Energy optimization.**

I Introduction

The advent of the Internet of Things (IoT) has heralded a new era in home living, where intelligence, connectivity, and efficiency converge to redefine the way we interact with our living spaces. This introduction embarks on a comprehensive exploration of the design and implementation of an IoT Smart Home Automation System, emphasizing the critical components of temperature and occupancy monitoring. As we delve into the intricacies of this cutting-edge system, the overarching goal is to understand how IoT technologies seamlessly integrate into the fabric of our homes, transforming them into intelligent, adaptive environments. The concept of a "smart home" has transcended its initial novelty to become a defining feature of modern living. As technology continues to advance, the integration of the Internet of Things (IoT) into home automation systems has given rise to intelligent, interconnected living spaces. This introduction embarks on an in-depth exploration of an IoT Smart Home Automation System, with a particular focus on temperature and occupancy monitoring. The evolution of smart homes is traced from the early days of home automation to the sophisticated and interconnected systems that characterize contemporary living.The roots of home automation can be traced back to the early 20th century, where visionary inventors and innovators experimented with mechanized devices to simplify household

ISSN: 2454-9924

tasks. The historical perspective highlights the gradual evolution from basic automation, such as timer-controlled lights, to more sophisticated systems enabled by advancements in computing and communication technologies. The advent of the Internet of Things marks a pivotal moment in the evolution of smart homes. This section delves into the fundamental principles of IoT and its integration into home automation. The interconnectedness of devices, facilitated by the IoT paradigm, lays the foundation for seamless communication and enhanced control over various aspects of the home environment.

The IoT Smart Home Automation System, with a focus on temperature and occupancy monitoring, is not merely a technological convenience but a transformative force shaping the way we inhabit and interact with our living spaces. From the historical roots of home automation to the dynamic and interconnected systems of today, the journey has been one of innovation, challenges, and immense potential. The integration of temperature and occupancy monitoring into the IoT ecosystem offers unprecedented control, comfort, and energy efficiency, laying the foundation for a future where homes are intelligent, responsive, and sustainable. As technology continues to evolve, the smart home becomes not just a collection of devices but a dynamic and adaptive environment that anticipates and meets the evolving needs of its inhabitants. This comprehensive exploration serves as a roadmap for understanding the past, present, and future of IoT Smart Home Automation, inviting us to envision a world where our living spaces are truly intelligent and harmonized with the rhythms of modern life.

II Related Work

The design and implementation of an IoT smart home automation system with temperature and occupancy monitoring is a multifaceted research area that encompasses a range of technological aspects. In the early stages of this literature survey[1], The Wikipedia page provided an overview of the fundamental concepts, applications, and historical developments within the realm of IoT, serving as a starting point for understanding the broader landscape of interconnected devices and systems. However, it is crucial to acknowledge the tertiary nature of Wikipedia as a source and recognize the necessity of supplementing this foundational knowledge with more specialized and scholarly literature. The subsequent phases of the literature survey will involve a deeper exploration of academic papers, conference proceedings, and reputable sources that specifically address the intricacies of integrating temperature and occupancy monitoring within IoT-based smart home systems.

This source[2] provides insights into the potential drawbacks and benefits of utilizing Bluetooth technology for home automation applications. Bluetooth is known for its short-range wireless communication capabilities, making it a viable option for smart home devices. The advantages may include ease of setup, low power consumption, and compatibility with a wide range of devices. However, this survey aims to critically assess the disadvantages highlighted in the source, shedding light on potential challenges such as limited range, interference issues, and security concerns.By incorporating this source into the literature survey, the goal is to gain a nuanced understanding of the role of Bluetooth in smart home automation, particularly its

ISSN: 2454-9924

relevance to temperature and occupancy monitoring. This information will contribute to a comprehensive examination of various communication technologies, aiding in the identification of optimal solutions for the design and implementation of an efficient and reliable IoT smart home automation system.

This source[3] likely delves into the utilization of GSM (Global System for Mobile Communications) technology in the context of home automation. GSM is renowned for its widespread use in mobile communications, and its integration into smart home systems presents unique opportunities and challenges. The literature survey will scrutinize the content of this paper to extract insights into how GSM technology can contribute to the enhancement of a smart home automation system, with specific attention to temperature and occupancy monitoring.

Potential themes covered in the survey may include the advantages and limitations of GSM-based automation, the efficiency of remote monitoring and control facilitated by GSM, and considerations for implementing temperature and occupancy sensors within this framework. By incorporating findings from this source into the broader literature survey, a more comprehensive understanding of the diverse technologies available for IoT-based smart home systems will be achieved, aiding in informed decision-making during the design and implementation phases.

This source[4] explores the application of GSM technology in conjunction with Arduino for home automation. Arduino, a popular open-source hardware and software platform, is frequently used in IoT projects, including smart home systems. The literature survey will delve into the specifics of how GSM and Arduino are integrated to facilitate home automation, with a particular focus on temperature and occupancy monitoring. Key aspects to be considered in the literature survey include the technical details of the GSM-based system, the role of Arduino in sensor integration and data processing, and the practical implications for monitoring and controlling temperature and occupancy within a smart home environment.

This paper [5] explores the application of ZigBee technology in the design of smart home systems. ZigBee, a low-power wireless communication standard, is commonly used in IoT applications, including smart homes. The literature survey will examine this source to gain insights into how ZigBee technology is leveraged to facilitate communication and control within a smart home environment, with a specific focus on temperature and occupancy monitoring. Key aspects to be considered in the literature survey include the technical details of the ZigBee-based smart home system, the integration of sensors for temperature and occupancy monitoring, and the overall efficiency and effectiveness of the system in providing real-time data and control capabilities.

This paper [6] delves into the integration of RFID (Radio-Frequency Identification) technology in the context of home automation, specifically focusing on energy metering and reporting. The literature survey will analyse this source to gather insights into how RFID technology is utilized to enhance energy management and automate various aspects of a smart home, potentially including temperature and occupancy monitoring.Key aspects to be explored in the literature survey include the technical details of the RFID-based system, the incorporation

ISSN: 2454-9924

of sensors for temperature and occupancy monitoring, and the overall effectiveness of the system in providing energy-efficient and automated functionalities. By incorporating findings from this source into the broader literature survey, a deeper understanding of the practical implementation of RFID-based smart home systems and their potential for temperature and occupancy monitoring will be achieved. This insight will contribute to the overall assessment of technologies available for IoT-based smart home systems, aiding in the informed design and implementation of a comprehensive solution.

This paper [7]explores the integration of RFID (Radio-Frequency Identification) technology in the context of smart home systems, with a focus on mobile applications and IoT services. The literature survey will analyse this source to gather insights into how RFID technology is used in conjunction with mobile devices to create an Internet-of-Things (IoT) ecosystem for smart homes, potentially encompassing temperature and occupancy monitoring.

Key aspects to be explored in the literature survey include the technical details of the RFID-based IoT system, the role of mobile devices in controlling and monitoring smart home functionalities, and the overall effectiveness of the system in providing seamless and user-friendly IoT services.By incorporating findings from this source into the broader literature survey, a deeper understanding of the practical implementation of RFID-based mobile IoT systems for smart homes and their potential for temperature and occupancy monitoring will be achieved. This insight will contribute to the overall assessment of technologies available for IoT-based smart home systems, aiding in the informed design and implementation of a comprehensive solution.

This paper [8] investigates the application of Wi-Fi-based wireless sensor networks in the context of low-cost home automation systems, incorporating IoT principles. The literature survey will analyze this source to gain insights into how Wi-Fi technology, coupled with wireless sensor networks, contributes to the creation of an IoT-enabled smart home, with a specific focus on temperature and occupancy monitoring.Key aspects to be explored in the literature survey include the technical details of the Wi-Fi-based wireless sensor network, the integration of sensors for temperature and occupancy monitoring, and the overall cost-effectiveness and efficiency of the system in providing IoT-enabled home automation services.By incorporating findings from this source into the broader literature survey, a deeper understanding of the practical implementation of low-cost, Wi-Fi-based smart home systems and their potential for temperature and occupancy monitoring will be achieved. This insight will contribute to the overall assessment of technologies available for IoT-based smart home systems, aiding in the informed design and implementation of a comprehensive solution.

This paper [9] proposed the utilization of speech-based control in a home automation system, integrating both Bluetooth and GSM technologies. The literature survey will analyse this source to gain insights into how speech recognition interfaces with Bluetooth and GSM for controlling and monitoring smart home functionalities, with specific consideration for temperature and occupancy.Key aspects to be explored in the literature survey include the technical details of the speech-based home automation

ISSN: 2454-9924

system, the integration of Bluetooth and GSM technologies, and the effectiveness of the system in providing a user-friendly and hands-free interface for smart home control and monitoring.By incorporating findings from this source into the broader literature survey, a deeper understanding of the practical implementation of speech-based smart home systems using Bluetooth and GSM, and their potential for temperature and occupancy monitoring, will be achieved. This insight will contribute to the overall assessment of technologies available for IoT-based smart home systems, aiding in the informed design and implementation of a comprehensive solution.

As part of the literature survey for the design and implementation of an IoT smart home automation system with temperature and occupancy monitoring, a relevant reference is the paper authored by Muhammad Asadullah and Ahsan Raza titled "An Overview of Home Automation Systems."

This paper[10] likely provides a comprehensive review of home automation systems, offering insights into the different technologies, protocols, and features employed in smart homes. The literature survey will analyse this source to gain a broad understanding of the landscape of home automation, exploring various aspects that may include temperature and occupancy monitoring. Key aspects to be explored in the literature survey include the overview of existing home automation technologies, the integration of sensors for temperature and occupancy monitoring, and the overall architectural considerations for building efficient and user-friendly smart home systems. By incorporating findings from this source into the broader literature survey, a foundational understanding of home automation technologies and their potential applications for temperature and occupancy monitoring will be achieved. This insight will contribute to the overall assessment of technologies available for IoT-based smart home systems, aiding in the informed design and implementation of a comprehensive solution.

The related works section provides a thorough examination of the existing landscape in IoT-based smart home automation systems, emphasizing temperature and occupancy monitoring. The multifaceted exploration of commercial solutions, residential implementations, research contributions, sensor technologies, interdisciplinary approaches, challenges, case studies, and future trends offers a holistic understanding of the current state and future potential of smart homes. This groundwork sets the stage for the subsequent design and implementation of an innovative and user-centric IoT Smart Home Automation System that addresses the evolving needs and challenges of modern living. The related works section unfolds as a comprehensive journey through the existing body of research, implementations, and innovations in the domain of IoT-based smart home automation systems. This exploration focuses particularly on the critical aspects of temperature and occupancy monitoring, delving into the rich tapestry of endeavours that have shaped the smart home landscape.

Commercially available smart home automation systems are investigated to discern prevalent trends, features, and user experiences. Notable systems such as Amazon's Alexa, Google's Home, and Apple's HomeKit are scrutinized for their functionalities and integrations. Case studies of residential smart home implementations

ISSN: 2454-9924

provide valuable insights into the diverse approach's homeowners take to integrate automation technologies. These studies serve as real-world benchmarks for understanding the practical applications and challenges faced in residential environments. A review of academic contributions in the realm of smart home automation focuses on advancements in IoT technologies, system architectures, and human-computer interaction. Notable research papers and projects are analysed to extract key findings and methodologies. This subheading explores experimental prototypes developed in academic settings that push the boundaries of smart home capabilities. These prototypes often serve as testbeds for novel ideas, showcasing the potential for future innovations. The landscape of temperature sensing technologies is explored, ranging from traditional thermocouples to more advanced and accurate solutions. Emphasis is placed on the importance of reliable temperature data for effective climate control. The subheading delves into the precision and accuracy of temperature sensors utilized in smart home systems, highlighting the significance of these attributes in ensuring optimal climate control. The integration of temperature monitoring systems with Heating, Ventilation, and Air Conditioning (HVAC) systems is examined. This exploration provides insights into how smart homes optimize energy efficiency through intelligent climate control. The implementation of dynamic climate control strategies based on real-time temperature data is discussed. This subheading explores how these strategies ensure both comfort and energy savings, a crucial aspect of smart home design.

This subheading evaluates the effectiveness of motion sensors and infrared technologies in accurately detecting occupancy. Their applications in smart home systems are discussed. Smart Cameras and Computer Vision: Advancements in smart camera technologies and computer vision algorithms are explored for more sophisticated occupancy monitoring. This includes applications in security and energy efficiency within smart homes. This section delves into how occupancy monitoring contributes to home security through the implementation of intrusion detection systems. It sheds light on the role of occupancy data in ensuring the safety of smart homes. The utilization of occupancy data to implement energy-saving strategies, such as intelligent lighting and HVAC control, is examined. This exploration highlights the dual role of occupancy monitoring in enhancing both security and energy efficiency.

Collaborations between IoT experts and architects are discussed, focusing on how smart home technologies seamlessly integrate into the architectural design of modern homes. The intersection of IoT and energy-efficient home designs is explored, showcasing interdisciplinary research aimed at creating homes that leverage IoT for optimal energy efficiency and occupant comfort. This subheading discusses studies focused on user interactions with smart home interfaces, emphasizing the significance of intuitive and user-friendly designs. Exploration of research on user feedback and adoption patterns sheds light on the factors influencing the acceptance of smart home technologies. This section delves into the importance of user-centric design in ensuring successful implementations.The challenges of privacy and security in IoT-based smart home

ISSN: 2454-9924

systems are dissected, with a focus on the implementation of data encryption and privacy measures.

The importance of user education in mitigating privacy and security concerns is discussed, acknowledging the role of informed users in maintaining the security of their smart homes.

Challenges arising from the diversity of devices and communication protocols within the smart home ecosystem are examined. This subheading highlights the complexities associated with ensuring seamless interoperability. Ongoing standardization efforts aimed at promoting interoperability among smart home devices are explored. The importance of establishing industry standards to create a seamless user experience is emphasized. Case studies of IoT-based smart home implementations in single-family residences are analysed. Successes, challenges, and user experiences provide valuable insights for future implementations. This subheading explores how smart home technologies are adapted for multi-family dwellings. Considerations of scalability and user diversity are discussed, offering lessons learned from diverse residential settings. The integration of smart home technologies in commercial office spaces is investigated, focusing on how these technologies enhance efficiency and occupant comfort. This section explores industrial applications of smart home automation principles for improved safety and operational efficiency. Real-world implementations in industrial settings are examined for their impact on processes and outcomes.

III PROPOSED METHOD

The proposed Smart Home Automation System comprises temperature sensors strategically placed to monitor and regulate the temperature within the home environment. When the temperature exceeds a predefined threshold, the system activates fans to maintain a comfortable atmosphere. Simultaneously, an IR sensor positioned at the entrance detects the presence of individuals entering the room, triggering the automatic illumination of lights. This not only enhances security by lighting up the space but also adds convenience for occupants. The incorporation of a Wi-Fi module enables real-time data transmission to the Thing Speak website. Users can remotely monitor temperature variations and occupancy status, allowing for proactive adjustments and energy management. The Thing Speak platform provides a user-friendly interface for data visualization and analysis. The proposed system aims to contribute to the development of energy-efficient and intelligent homes, where occupants can enjoy an enhanced living experience while minimizing energy consumption. The system can optimize heating, ventilation, and air conditioning (HVAC) systems based on real-time temperature data, leading to energy savings. Lights, heating, and cooling systems can be automatically adjusted based on occupancy, reducing energy consumption when rooms are unoccupied. Users can remotely control and monitor their home environment through smartphones or other devices. Automation of routine tasks, such as adjusting thermostats or turning on/off lights, enhances convenience. Occupancy monitoring can be integrated into security systems, providing alerts for unexpected movements or intrusions. Smart locks and surveillance systems can be integrated for enhanced home security. Collecting and

ISSN: 2454-9924

analysing data over time can provide insights into energy usage patterns and occupancy trends, enabling better decision-making for resource optimization. Integration with other IoT devices like smart speakers, cameras, and appliances can create a cohesive and interconnected smart home ecosystem. Users can monitor and control the system remotely, providing flexibility and peace of mind.

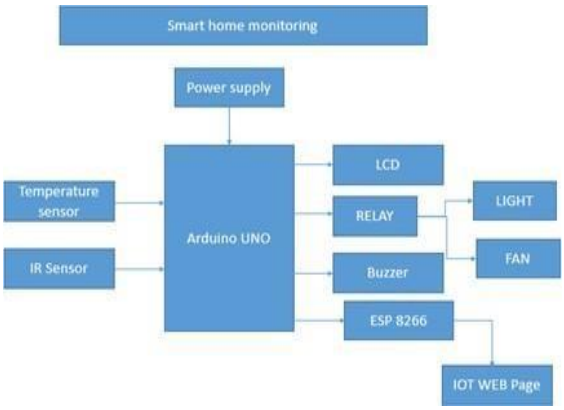


Fig 1: Block Diagram

The block diagram for the "Design and Implementation of IoT Smart Home Automation System with Temperature and Occupancy Monitoring" using a temperature sensor, IR sensor, LCD, relay, buzzer, ESP8266, fan, light, and Arduino Uno can be divided into several functional blocks, each representing a specific component or feature of the system. Below is a detailed explanation of the block diagram:

Monitors the ambient temperature in the environment. The temperature sensor is connected to the Arduino Uno to measure temperature data. *Function:* Detects the presence or absence of occupants in the room. The IR sensor is interfaced with Arduino Uno to provide occupancy data. Displays real-time information such as temperature and occupancy status. Connected to the Arduino Uno for receiving and displaying data. Controls high-power devices such as the fan and light. Connected to the Arduino Uno to receive control signals based on occupancy and temperature conditions.

Provides audible alerts or notifications. Connected to the Arduino Uno and triggered based on specific events such as intrusion detection or critical temperature levels. Enables communication with the IoT ecosystem for remote monitoring and control. Interfaced with Arduino Uno to transmit data to and receive commands from the IoT platform. Devices controlled based on temperature and occupancy conditions. Connected to relays, which are in turn controlled by the Arduino Uno based on the data from the temperature sensor and IR sensor. Serves as the central processing unit, collecting data from sensors, making decisions based on predefined conditions, and controlling actuators accordingly. Receives data from the temperature sensor and IR sensor. Controls the relay for the fan and light based on temperature and occupancy conditions. Sends data to the LCD display for local information. Interfaces with the ESP8266 for IoT communication. Facilitates remote monitoring and control of the smart home system. Communicates with the ESP8266, allowing users to access and control the system remotely through a web or mobile interface.

ISSN: 2454-9924

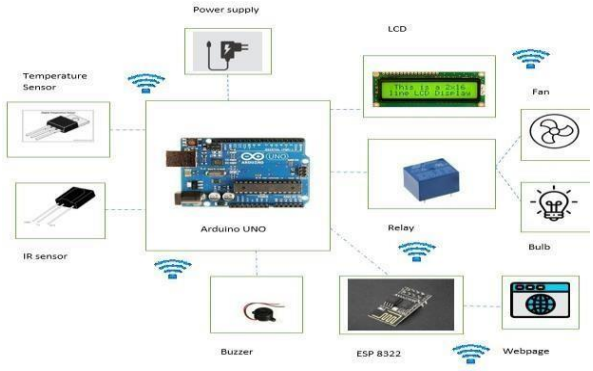


Fig:2 Communication diagram

The communication diagram for the design and implementation of an IoT smart home automation system with temperature and occupancy monitoring involves the interaction of various components to create a seamless and efficient home automation experience. At the core of the system are the temperature sensor, IR sensor, power supply, relay, ESP8266 (assuming there's a typo in "esp8322"), LCD display, Arduino microcontroller, and IoT connectivity. The temperature sensor continuously measures the ambient temperature, while the IR sensor detects occupancy in specific areas of the home. These sensors feed their data to the Arduino microcontroller, which serves as the brain of the system. The microcontroller processes the sensor data and makes decisions based on predefined rules and user preferences. For instance, if the temperature exceeds a certain threshold, the system may activate the HVAC (Heating, Ventilation, and Air Conditioning) system through the relay. The Arduino microcontroller is also responsible for controlling other connected devices, such as turning on or off lights or appliances based on occupancy detected by the IR sensor. The LCD display provides real-time feedback on temperature, occupancy status, and system actions, ensuring users are informed about the system's operations. The IoT connectivity, facilitated by the ESP8266 module, enables the smart home system to communicate with an external server or cloud platform. This connectivity allows users to remotely monitor and control their smart home through a dedicated mobile app or web interface. Users can receive temperature alerts, check occupancy status, and manually override system settings as needed. The power supply ensures that all components receive the necessary electrical power to operate smoothly. The communication diagram illustrates the data flow and interactions among these components, showcasing how the sensors, microcontroller, and connectivity modules collaborate to create an intelligent and responsive home automation system. Overall, this integrated system enhances energy efficiency, comfort, and security in the home by leveraging IoT technology for intelligent automation and remote management.

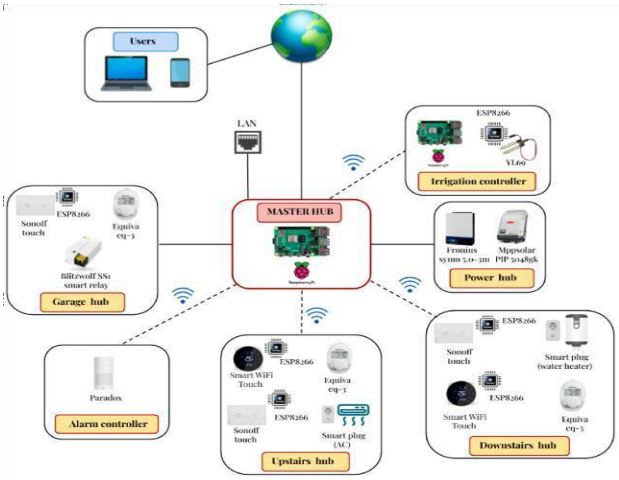


Fig.3: Architecture Diagram

The architecture diagram for the design and implementation of an IoT smart home automation system with temperature and occupancy monitoring is a comprehensive representation of the system's structure and the interactions among its key components.

ISSN: 2454-9924

At the heart of the architecture is the Arduino microcontroller, which serves as the central processing unit for the system. Connected to the Arduino are the temperature sensor and IR sensor, responsible for monitoring the environmental conditions within the home. The temperature sensor measures ambient temperature, while the IR sensor detects occupancy in various areas. The Arduino, acting as the brain of the system, processes the data from these sensors and makes decisions based on predefined rules and user preferences. The relay is employed to control devices such as the HVAC system or lights, ensuring responsive and automated adjustments to temperature and occupancy changes. The LCD display provides a user-friendly interface, offering real-time feedback on temperature status, occupancy, and system actions. The IoT connectivity is facilitated by the ESP8266 module, allowing the smart home system to communicate with external servers or cloud platforms. This connectivity enables remote monitoring and control, as users can access the system through a dedicated mobile app or web interface. The cloud platform plays a crucial role in processing incoming data, storing historical information, and facilitating communication between the smart home system and the user interface. The power supply ensures that all components receive the necessary electrical power to function properly. Altogether, this architecture diagram illustrates a cohesive and scalable system where the components collaborate seamlessly to create an intelligent, energy-efficient, and user-friendly IoT smart home automation system. The integration of sensors, microcontroller, connectivity modules, and display elements demonstrates a holistic approach to enhancing home automation, making it responsive to environmental changes and accessible to users both locally and remotely.

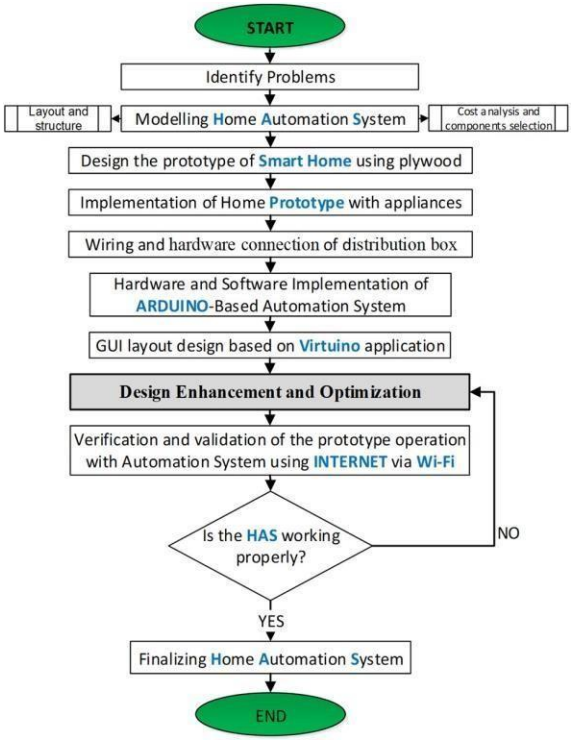


Fig.4: Flow Diagram

Start: The flow diagram begins with the start point. Initialize System: Initialize the smart home automation system, including powering up the components. Sensor Data Acquisition: The temperature sensor and IR sensor continuously monitor the environment. The temperature sensor measures the ambient temperature. The IR sensor detects occupancy in specific areas. Data Processing (Arduino): The Arduino microcontroller processes the sensor data. Decision-making based on predefined rules and user preferences: If the temperature is above a threshold, activate HVAC (Heating, Ventilation, and Air Conditioning) through the relay. Control lights or appliances based on occupancy detected by the IR sensor. Display Information (LCD): Display real-time information on the LCD screen:

ISSN: 2454-9924

Temperature status. Occupancy status. System actions or alerts. IoT Connectivity (ESP8266): The ESP8266 module facilitates communication with external servers or cloud platforms. Transmit sensor data and system status to the cloud. Cloud Processing: The cloud platform processes incoming data. Allows remote monitoring and control via a dedicated mobile app or web interface. User Interaction: Users can receive alerts. Check temperature and occupancy status remote End: The flow diagram concludes at the endpoint. This flow diagram provides a high-level overview of the sequence of actions in the IoT smart home automation system. It showcases how the sensors, Arduino microcontroller, LCD display, and IoT connectivity work together to monitor and control the home environment efficiently. Keep in mind that the actual flow may involve more detailed steps depending on the specific functionalities and features.

IV System Design

Temperature Monitoring

The temperature sensor continuously measures the ambient temperature. Arduino Uno reads temperature data and determines if it exceeds a predefined threshold. The IR sensor detects the presence or absence of occupants in the room. Arduino Uno receives occupancy data and adjusts the system's behaviouraccordingly. Arduino Uno sends temperature and occupancy status to the LCD display for local monitoring. Based on temperature and occupancy conditions, Arduino Uno controls the relays. The relay controls the fan and light, turning them on or off as needed. The buzzer is triggered for specific events, such as detecting an intruder or reaching critical temperature levels. ESP8266 communicates with the IoT platform, providing real-time data and receiving commands for remote monitoring and control. Users can monitor and control the smart home system remotely through the IoT platform. By integrating these components and functionalities, the system ensures a responsive, energy-efficient, and secure smart home environment with temperature and occupancy monitoring capabilities.

V Results and Discussion

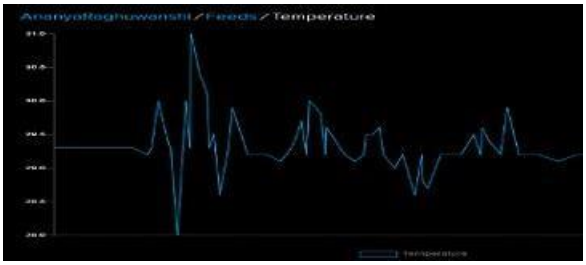


Figure 5. DHT 11 Temperature sensing graph

The above fig [5] line graph shows the real-time output of the DHT 11 Sensor. The sensor uploads the magnitude of changes in Temperature & Humidity in every 30 sec of interval. The fig [5] DHT11 graphs are in Temperature (Celsius) vs Time (sec) and Humidity (percentage) vs Time (sec) in every 30 second's change. The data changes respectively to the change in the place where the sensor is placed.

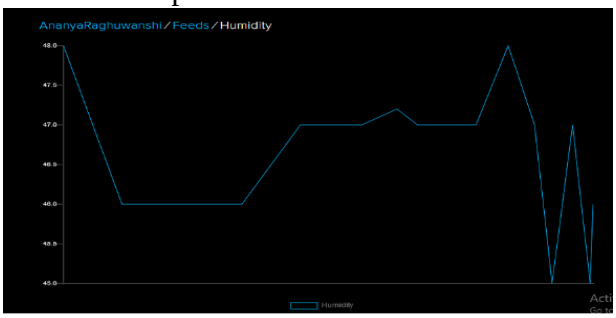


Figure 6. Sound intensity graph

ISSN: 2454-9924

The above line graph fig [6] shows the real-time output of the LM 358 module with microphone. The sensor uploads the magnitude of changes in the sound intensity in every 30 sec of interval.

Figure 7. Humidity sensing graph

The graph fig [7] shows the variation of the sound intensity in percentage with time. The data changes respectively to the change in the place where the sensor is placed.

Figure 8: Light Intensity graph

The above line graph fig [8] shows the real-time output of the PIR Sensor. The activity in the range (7m) of sensor is detected and the output is either zero or one depending on the level of activity. The data changes respectively to the change in the place where the sensor is placed.

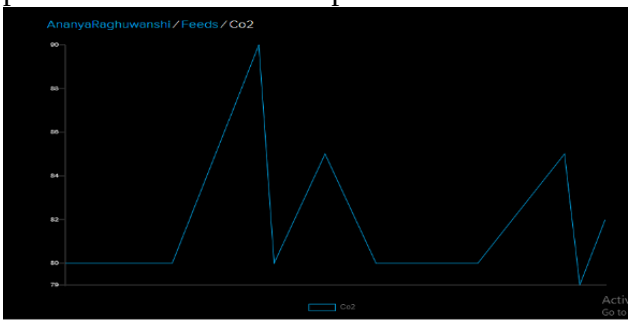


Figure 9: Co_2 variation graph

The above line graph fig [9] shows the real-time output of the MQ135 gas sensor module. The sensor uploads the magnitude of changes in the CO2 (in ppm) in every 30 sec of interval. The graph fig [9] shows the variation of the gas percentage with time. The data changes respectively to the change in the place where the sensor is placed.

VI Conclusion

In conclusion, the "Design and Implementation of Smart Home Automation System using IoT" represents a significant leap towards creating intelligent living spaces that seamlessly integrate cutting-edge technologies for enhanced comfort, energy efficiency, and security. Through the amalgamation of diverse components such as temperature sensors, IR sensors, LCD displays, relays, buzzers, ESP8266, and more, this system exemplifies a holistic approach to modern home automation. The integration of a temperature sensor enables precise climate control, ensuring optimal comfort and energy efficiency. The occupancy detection system, powered by IR sensors, not only enhances security by detecting intruders but also contributes to energy conservation through intelligent lighting and HVAC control. The Arduino Uno acts as the central intelligence hub, processing data from sensors, making decisions based on predefined conditions, and orchestrating the control of actuators. The inclusion of a user-friendly LCD display provides occupants with real-time information, fostering transparency and enabling manual control when desired. The relay system empowers the automation of high-power devices like fans and lights, responding dynamically to environmental changes and occupancy status. The ESP8266 facilitates seamless communication with IoT

ISSN: 2454-9924

platforms, extending the system's reach to remote monitoring and control, thereby aligning with the contemporary paradigm of connected living. The project's success lies not only in its technical functionalities but also in its emphasis on user-centric design. The implementation takes into account the diverse needs of occupants, offering a system that is intuitive, adaptable, and enhances the overall living experience. The inclusion of a buzzer for audible alerts adds an extra layer of security, notifying occupants of critical events such as intrusion or extreme temperatures. As technology continues to advance, the project provides a foundation for future innovations in smart home automation. The system's adaptability to emerging sensor technologies and its integration with IoT platforms pave the way for further enhancements. The machine-to-machine communication enabled by the IoT framework allows for the development of more sophisticated algorithms, predictive analytics, and personalized automation, making the smart home even more responsive to the needs and preferences of its inhabitants.

Vii References

[1] Varalakshmi, I., Thenmozhi, M. and Sasi, R., 2021, July. Detection of Distributed Denial of Service Attack in an Internet of Things Environment-A Review. In 2021 International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-6). IEEE.

[2] Bluetooth based automation advantages and disadvantages,24.07.2018, https://www.quora.com/What-are-the-disadvantages-ofusing-home-automation-via-Bluetooth

[3] GSM based automation, 24.07.2018, retrieved fromhttps://www.slideshare.net/MainakSinha1/gsm-based-homeautomation-62745555.

[4] GSM based automation, 24.07.2018, retrieved fromhttps://circuitdigest.com/microcontroller-projects/gsm-based-homeautomation-using-arduino

[5] GSM based automation, 24.07.2018, retrieved fromhttps://www.slideshare.net/MainakSinha1/gsm-based-homeautomation-62745555.

[6] Varalakshmi, I., M. Thenmozhi, and R. Sasi. "Detection of Distributed Denial of Service Attack in an Internet of Things Environment-A Review." 2021 International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, 2021.

[7] Varalakshmi, I., & Kumarakrishnan, S. (2019, March). Navigation system for the visually challenged using Internet of Things. In 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-4). IEEE.

[8] Mohsen Dardanian, Martin Peter Michael, “Smart Home Mobile FID Based Internet-of-Things Systems and Services," International Conference on Advanced Computer Theory and Engineering, 2008

[9] N. Vikram, K.S. Harish, M.S. Nihaal, Raksha Umesh, Aashik Shetty, Ashok Kumar, “A Low-Cost Home Automation System Using Wi-Fi Based Wireless Sensor Network Incorporating Internet of Things(IoT)," 7th International Advance Computing Conference,2017.

[10] Dhiraj Sunera, Vemula Tejaswi, “Implementation of Speech Based Home Automation System Using Bluetooth and GSM," International Conference on Signal Processing, Communication, Power and EmbeddedSystem (SCOPES), 2016.

ISSN: 2454-9924

[11] Muhammad Asadullah, Ahsan Raza , An Overview of Home Automation Systems
[12] Varalakshmi, M.I. and Thenmozhi, M., 2021. Mitigation of DDoS attack using machine learning algorithms in SDN_IoT environment. Design Engineering, pp.4381-4390.
[13] Paul Jasmin Rani, Jason Bathukamma, B. Praveen Kumaar, Gurpraveen Kumaar, Santhosh Kumar, “Voice-controlled home automation system using Natural Language Processing (NLP) and Internet of Things (IoT)," Third International Conference on Science Technology Engineering & Management (ICONSTEM), 2017.
[14] S. M. Brundha, P. Lakshmi, S. Santhanalakshm, “Home automation in client-server approach with user notification along with efficient security alerting system," International Conference on SmartTechnologies for Smart Nation (Matichon), 2017.
[15] Varalakshmi, I., et al. "Smart Dumpster Monitoring System Using Efficient Route-Finding Algorithm." 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, 2019.

ICADET: conference proceedings: 2024 Advanced Development in Engineering And Technology

ISSN: 2454-9924

AN ADVANCED MOBILE TRACKING SYSTEM WITH PIN POINTS USING ANDROID SMART PHONES

Dr.K.Venkata Rao[#1] and K.Vineet[*2]

[#1]*Associate Professor, Department of Computer Science & Systems Engineering, Andhra University, Visakhapatnam, India*

[*2]*M.Tech (C.S.T), , Department of Computer Science & Systems Engineering, Andhra University, Visakhapatnam, India.*

Abstract*— : *Smart Phones has surely become one of the valuable gadgets for human beings. It is necessary for us to have a dependable device which will provide all the facilities other than the basic functionalities available in a mobile phone. This project involves an Android Application Development of a Personalized GPS based Location Tracker in which any Android mobile device (app installed) can locate any other GPS enabled handset (app installed). Though target user may be located anywhere in the world, he must have network connectivity, provided GPS enabled. Personal Tracking Systems are the devices specially built up for personal safety. This tracking system has been implemented, which adopts various extended features that the existing system does not have. Our application provides the functionality in which user's safety is ensured.

***Keywords* — Tracking, Global Positioning System (GPS), Global System for Mobile Communication (GSM), Android, Mobile Application, tracking using a Smartphone**

I Introduction

Android is a mobile OS (Operating System) based on Linux Kernel and currently developed by Google, with a user interface based on direct manipulation. Android provides a rich application framework that allows you to build innovative apps and games for mobile devices in a Java language environment. Android apps are built as a combination of distinct components that can be invoked individually. Turning the GPS module on the phone would not cost us anything but getting a location usually involves transaction with cell phone service provider so as to extract the location fast and with as little network connectivity as possible plus non visibility of satellites. Able to install custom apps from the market, GPS, Location through the network, use all social media sites and e-commerce sites through their apps are most popular features.

There are many companies where they have a need to track their employees periodically throughout the day reasons being to avoid employee cheating the employer by not visiting the places he has been asked to or to track employee performance by real-time data or showing miscellaneous expenditure without actually spending or using it example, travelling charges. Earlier GPS systems uses only single user tracking environment to track the location. The main disadvantage of these earlier systems is that it takes a large amount of time to track the exact location of the user. In this paper, we are proposing an application using Andriod environment which is used to track two users and also for every 30 sec it tracks the exact pin-point location of the user continuously until our application is turned off.

ISSN: 2454-9924

II Related Work

Most applications in the market are not user friendly because they do not provide precise data, nor allow multiple ways to access the data. Currently GPS tracking system using android Smartphone‟s only locates maps with reference to longitude and latitudes. This technique only gives location details. So it is not an easy task to find exact location. The proposed system is meant to resolve such deficiencies. It uses the cell phone service provider to locate the requester for a registered service. Present technique differs from many other types of mobile services because it is not just mobile in the sense that it can be carried with the user but it can actually be used on the move . In addition, it takes into consideration the usage situations that may affect the location‟s physical environment (e.g., background noise, illumination, weather).It also takes a lot of time to track the moving object which has resolved in the proposed system.

III MATHEMATICAL DESCRIPTION

PSEUDOCODE:

Input: radius of the sphere(r), longitudes long1, long2 of both the points p1 and p2, latitudes lat1, lat2 of both the points p1 and p2.

Output: distance(d) between two points p1 and p2.

Method:

1. Calculate r = radius of the sphere
1. Calculate lat1 = latitude of point 1
2. Calculate lat2 = latitude of point 2
3. Calculate long1 = latitude of point 1
4. Calculate long2 = latitude of point 2
5. Calculate a = Sin^2((lat2-lat1)/2) + Cos(lat1)Cos(lat2) * Sin^2((long2-long1)/2)
6. Calculate c= 2 * a * tan2(Sqrt(a) * Sqrt(1-a))
7. Distance d = r * c

IV System Design

An Advanced Mobile Tracking System with Pin Points Using Android Smart Phones introduces the architecture and component models of Android, and analyzes the anatomy of an Android application including the functions of Activity, Intent Receiver, Service, Content Provider, and etc., The method of a location-based mobile service is implemented using Android. This design example shows that its much effortless to implement self-location, to trace the user‟s location, to perform query and to flexibly control the real-time map on Android.

Here in this Android Application Development of a Personalize GPS based Location Tracker in which any Android mobile device (app installed) can locate any other GPS enabled handset (app installed). Though target user may be located anywhere in the world, he must have network connectivity, provided GPS enabled.

ISSN: 2454-9924

Screen 1

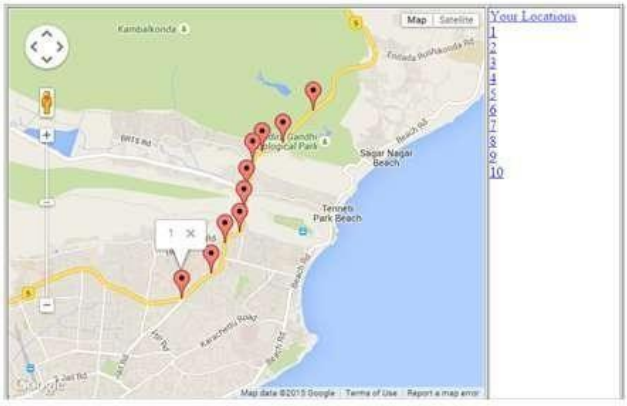
Providing the users IP address in the URL and even providing the server URL, will help in pushing the user location data to server till the GPS is disabled.

Screen 2

Once the tracker is enabled in the server side it locates the user device in a map.

Screen 3

This screen shows the pinpoint locations for every 30 seconds of the user who is moving until the GPS is off.



Screen 4

Here in the Mobile Tracker on providing the second user i.e(Friend‟s Mail Address), would track his details even.

Screen 5

The obtained pin point locations on the map are been traced, which are obtained by the perodic updated locations of the devices

ISSN: 2454-9924

from the servers information.

This pin pointed locations are been trajected till the GPS is disabled.

Screen 6

Once the details are been provided, the locations of the individual users are been traced on a map. Thus we achived the pin point locations of these devices.

Screen 7

VI Algorithm/ Techniques Description

HAVERSIN ALGORITHM

This Algorithm is used for calculating distance between two Geographical Locations. When you want to calculate the distance between the locations, you cannot head towards the east direction in a straight line, where you will be stopped before 30 meters from the destination point because the earth is spherical and it follows different Geographical Poles.(The Northern direction is always the shortest route). Hence, we always calculate with 90° because it‟s navigation is on North Side and it goes in a clockwise direction as positive.

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

$$haversin\left(\frac{d}{r}\right) = haversin(\phi_2 - \phi_1) + c$$

Where haversin is the heversine function

$$haversin(\theta) = \sin^2\left(\frac{\theta}{2}\right) = \frac{1-\cos(\theta)}{2}$$

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

On the left side of the equals sign d/r is the central angle, assuming angles are measured in radians (note that φ and λ can be converted from degrees to radians by multiplying by π/180 as usual).Solve for „d‟ by applying the inverse haversine function:

$$d = r\,\text{haversin}^{-1}(h) = 2r\arcsin\left(\sqrt{h}\right)$$

Where h is haversin (d/r) , or more explicitly:

$$d = 2r\arcsin\left(\sqrt{haversin(f_2 - f_1) + \cos(f_2)haversin(\lambda_2 - \lambda_1)}\right)$$

$$= 2r\arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

4.The law of haversines :

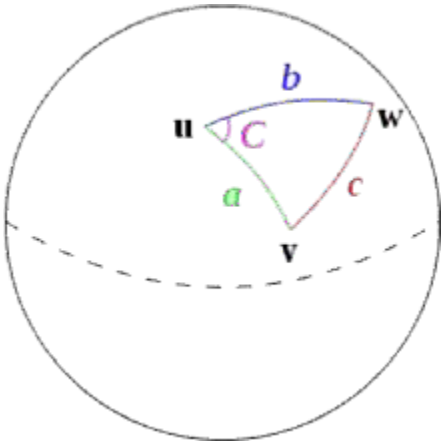
ISSN: 2454-9924

Given a unit sphere, a "triangle" on the surface of the sphere is defined by the great circles connecting three points **u**, **v**, and **w** on the sphere. If the lengths of these three sides are *a* (from **u** to **v**), *b* (from **u** to **w**), and *c* (from **v** to **w**), and the angle of the corner opposite *c* is *C*, then the law of haversines states:

(the law of haversines)

$$\operatorname{haversin}(c) = \operatorname{haversin}(a-b) + \sin(a)\sin(b)\operatorname{haversin}(C)$$

Since this is a unit sphere, the lengths *a*, *b*, and *c* are simply equal to the angles (in radians) subtended by those sides from the center of the sphere (for a non-unit sphere, each of these arc lengths is equal to its central angle multiplied by the radius of the sphere).



Spherical triangle solved by the law of haversines.

In order to obtain the haversine formula of the previous section from this law, one simply considers the special case where **u** is the north pole, while **v** and **w** are the two points whose separation *d* is to be determined. In that case, *a* and *b* are $\pi/2 - \varphi_{1,2}$ (i.e., 90° − latitude), *C* is the longitude separation $\Delta\lambda$, and *c* is the desired d/R. Noting that $\sin(\pi/2 - \varphi) = \cos(\varphi)$, the haversine formula immediately follows.To derive the law of haversines, one starts with the spherical law of cosines:

As mentioned above, this formula is an ill-conditioned way of solving for *c* when *c* is small. Instead, we substitute the identity that $\cos(\theta) = 1 - 2\operatorname{haversin}(\theta)$, and also employ the addition identity $\cos(a - b) = \cos(a)\cos(b) + \sin(a)\sin(b)$, to obtain the law of haversines, above.

VII Methodology Description

With modern technology, it‟s now possible to do many things on mobile phones and smart phones. Apart from the obvious convenience of being able to call colleagues and friends whilst on the move, smart phones can also be vital tools for use in business and commerce. But did you know that you‟re smart phone‟s built-in GPS receiver can also help you and your loved ones stay safe, by avoid them getting lost or find your way to that crucial meeting on time? By using a combination of GPS data, your current location can be established wherever your phone is capable of receiving a signal (even in hazardous climates). So, is this a good thing to make sure we are safe . Phone-tracking can be useful in both business and private life. Therefore my aim is to develop a mobile tracking application which is advantageous to know the precise location of an employee or family member at any given time.

VIII Conclusion

The localization on GPS is difficult when the objects are indoor or sheltered from the buildings and trees. The localization on wireless networks only has low accuracy and work in the situation of a disaster like the earthquake. A localization combined with GPS and wireless networks is built to make sure the consumers can expect greater safety and high accuracy location based services. It is a step towards pervasive positioning service. A J2ME mobile application based on providing Location Based Service using Global Positioning System (GPS) as a location provider is presented. The application is aware of the user with his current location coordinates and shows it on Google Maps. The application is also implemented as a client server system that helps users to locate their friends or anyone with whom he wants to share his location. The average location accuracy using this system is believed to be within a couple of meters. The application works in open space areas only since it relies on GPS. Future extensions may look at other options such as getting the location from the service provider. In this case the location accuracy will be reduced and will depend on the size of the cells where the user is located.

Other future extensions can be summarized as follows:

- Improvement in user Interfaces.
- Supports Multi-users to track their location.
- Support for external Bluetooth GPS receiver.
- Accuracy can be improved by several algorithms

IX References

[1] Ankur Chandra, GPS Locator: An application for Location Tracking and sharing using GPS for Java Enabled Handhelds, International Conference on Computational Intelligence and Communication Systems, 406-410, 2011.

[2] Litao He, Location Based Services combines with GPS and 3G Wireless Networks, IEEE, 542-545, 2008.

[3] Herwig MAYR, I-Navigate: Intelligent, Self-adapting Navigation Maps, IEEE, 0-7695-2772-8, 2007.

[4] Qia Wang, Video Based Real-World Remote Target Tracking On Smart Phones, IEEE International Conference on Multimedia and Expo, 693-698, 2012

[5] Xianhua Shu, Zhenjun Du, Rong Chen ; “Research on Mobile Location Service Design Based on Android” 978-1-4244-3693-4/09/ ©2009 IEEE.

[6] Ghaith Bader Al-Suwaidi; Mohamed Jamal Zemerly; “Locating Friends and Family Using Mobile Phones With Global Positioning System (GPS)” in Computer Systems and Applications, 2009, AICCSA 2009. IEEE/ACS International Conference on 10-13 May 2009

[7] Axel Küpper , Location-based services, fundamentals and operation, WILEY, 2nd edition, 2005.

[8] Jami, I.; Ali, M.; Ormondroyd, R.F.; "Comparison of methods of locating and tracking cellular mobiles" in Novel Methods of Location and Tracking of Cellular Mobiles and Their System Applications (Ref. No. 1999/046), IEE Colloquium on 17 May.

ICADET: conference proceedings: 2024 Advanced Development in Engineering And Technology
ISSN: 2454-9924

An efficient online image retrieval based on user query expansion

Ms. N. REKHA M.E, (P.G-Scholar)
Department of Computer Science and Engineering,
Arunai Engineering College,
Tiruvannamalai, India.
anjali.reks19@gmail.com

Mr. P. JAYAKUMAR Assistant Professor,
Department of Computer Science and Engineering,
Arunai Engineering College,
Tiruvannamalai, India.
jai13it@gmail.com

Abstract—Image re-ranking is a valuable method for an online-based image search. The examine based on only keywords pressed by the users is not proficient and results in unfixed output. The online-based image search recycled by Bing and Google uses image re-ranking technique. In an image that, users' objective is caught by one-click on the query image. This supports in given that better search results towards the users. Now we evaluate the technique in which a query keyword is first recycled to get back an excess of images constructed on the keyword. Image re-ranking structure mechanically learns dissimilar semantic spaces offline for dissimilar query keywords. Their visual structures are projected into their associated semantic spaces to catch semantic signatures for images. Images are re-ranked by differentiating their semantic signatures and the query keyword throughout the wired stage. The query-specific semantic signatures, meaningfully increase both the accuracy and efficiency of the re-ranking procedure. In future, it is proved to be a better method than the conservative online-based image search techniques.

Keywords: Re-ranking, query image, query keyword, semantic signature.

I. INTRODUCTION

Web image search engines use keywords as queries and search images based on the text associated with them. It is difficult for users to accurately describe the visual content of target images only, using keywords and hence text-based image search suffers from the uncertainty of query keywords. For instance, consuming apple as a query keyword, the regained images fit to dissimilar categories, such as apple laptop, apple logo, and apple fruit. To capture users' search intention, additional information has to be used in order to solve the ambiguity. Text-based keyword expansion is one way to make the Textual description of the query more detailed. Existing methods find either synonyms or other linguistics-related words from the thesaurus. However, the intention of users can be highly diverse and cannot be accurately captured by these expansions, even with the same query keywords. Content-based image retrieval with relevance feedback is widely used in order to solve this ambiguity. Users are required to select multiple relevant and irrelevant image examples and the visual similarity metrics are learned through online training from them. Images are re-ranked based on the learned visual similarities. However, for web-scale commercial systems, user's response has to be incomplete to the least lacking online training. In the method reviewed in this paper, a query keyword is first recycled to regain a set of images created on the keyword. Then the user is asked to pick an image from these images. Also, the rest of the images are ranked based on their visual similarities. The major challenge is the correlation of similarities of visual features and images' semantic meaning, which are needed to interpret users' intention to search. Recently, it has been suggested to contest images in a semantic space that used attributes or reference classes closely associated to the semantic meanings of images as base. Conversely, characterizing the highly varied images from the network is challenging because it is impossible to learn a universal visual semantic space. In this, a new framework was proposed to re rank the web images. As a substitute of physically describing a universal conception dictionary, it studies dissimilar semantic spaces for dissimilar query keywords independently and mechanically. The semantic space associated to the

images to be re ranked and it can be meaningfully pointed down by the query keyword delivered through the user. For instance, if the query keyword is "Paris," the concepts of "mountain" and "apple" are irrelevant to the query keyword so it will be excluded. As a replacement for, the concept of"computer" and "fruit" will be used as magnitudes to acquire the semantic space associated to the query keyword "apple". The semantic correlation among perceptions are explored and combined while calculating the resemblance of semantic signatures. Another important issue in this paper is, we didn't consider increasing the miscellany of search result by eliminating near-duplicate or much related images. The query specific semantic signatures are suggested to diminish semantic gap on the other hand it can't openly raise the miscellany of search result.

II. RELATED WORK

The fundamental factor of image re-ranking is to calculate visual similarities replicating semantic significance of images. Many visual features have been established in recent years. Though, instead of dissimilar query images, the active low-level visual features are diverse. Consequently, query images are classified into eight predefined significance arrangements and gave dissimilar feature weighting schemes to diverse types of query images. But to cover the large diversity of all the web images it was difficult for the eight weighting schemes. It was also probably for a query image to be categorized to a wrong classification. In order to decrease the semantic gap, query-specific semantic signature was first proposed and it recently increased each image with related semantic features over propagation in excess of a visual graph and a textual graph which were correlated. Alternative way of learning visual resemblances without tallying users' liability is pseudo relevance feedback. It takes the top N images most visually related to the query image by means of stretched optimistic examples to learn a resemblance metric. Computing the visual similarities that reflect the semantic relevance of images is the key component of image re-ranking. Many visual features have been developed in recent years. However, the operative low-level visual features are dissimilar for different query images. Therefore, Cui *et al*. categorized query images into eight predefined objective categories and offered diverse feature weighting schemes to dissimilar types of query images. But it was difficult for the eight weighting schemes to cover the large diversity of all the network images. It was also expected for a query image to be categorized to a wrong category. Query-specific semantic signature was first proposed in in order to reduce the semantic gap.

There is a lot of work on using visual features to re-rank images retrieved by initial text-only search, however, without requiring users to select query

images. Jing and Baluja proposed Visual Rank to analyze the visual link structures of images and tofind the visual themes for re-ranking. Cai*et al.* re-ranked images with attributes which were manually defined and learned from manually labeled *Harshil Jain et al* training samples. These approaches assumed that there was one major semantic category under a query keyword. Images were re-ranked by modeling this dominant category with visual and textual features.

A. Re-Ranking without Query Images

Query-specific semantic signature can be applied to image re-ranking without selecting query images. This application also requires the user to input a query keyword. But it assumes that images returned by initial text-only search have a dominant topic and images belonging to that topic should have higher ranks. Existing approaches typically address two issues: (1) how to compute the similarities between images and reduce the semantic gap; and (2) how to find the dominant topic with ranking algorithms based on the similarities. The query-specific semantic signature is effective in this application since it can improve the similarity measurement of images.

The query-specific semantic signature is also effective in this application, where it is crucial to reduce the semantic gap when computing the similarities of images. Due to the ambiguity of query keywords, there may be multiple semantic categories under one keyword query. These approaches cannot accurately capture users' search intention without query images selected by users. In recent times, general image appreciation and toning, there have been a amount of works on using projections over predefined concepts, attributes or reference classes as image signatures. The classifiers of concepts, attributes, and reference classes are trained from known classes with labeled examples. But the knowledge learned from the known classes can be transferred to recognize samples of novel classes which have few or even no training samples. Since these concepts, attributes, and reference classes are defined with semantic meanings, the projections over them can well capture the semantic meanings of new images even without further training. Rasiwasia *et al* plotted visual features to a worldwide concept vocabulary for image retrieval. Attributes with semantic meanings were used for object detection and recognition, face recognition, action recognition, image search and 3D object retrieval. Lampert *et al.* predefined a set of attributes on an animal database and detected target objects based on a combination of human-specified attributes instead of training images. Parikh and Grauman proposed relative attributes to indicate the strength of an attribute in an image with respect to other images. Some methods transportedinformation between object classes by calculating the similarities amongst novel object

classes and known object classes are called reference

classes. For example, Torresani*et al.* proposed an image descriptor which was the output of a number of classifiers on a set of known image classes, and used it to match images of other unrelated visual classes.

Online image re-ranking limits users' effort to just one-click feedback is an effective way to improve search results and its query not only increase the computational cost but also deteriorate the accuracy of re-ranking. However, how to find such relevant concepts automatically and use them for online web image re-ranking was not well explored in the conventional.

III. PROPOSED SYSTEM

The new image re-ranking framework focusses on the semantic signatures associated with the images. These semantic signatures are derived from the visual features associated with the images but are much shorter than the visual features. The diagram of the approach is shown in Fig. 2. It has offline and online portions. At the offline point, the reference classes (which represent different concepts) related to query keywords are automatically discovered and their training images are automatically collected in several steps. For a query keyword (for example apple), automatic selection of a set of maximum related keyword expansions (such as red apple and apple MacBook) is performed utilizing both textual as well as visual information. This set of keyword expansions describes the reference classes for the query keyword. In order to robotically acquire the training instances of a reference class, the keyword expansion (e.g., red apple) is recycled to regain images by the search engine originated on textual information over again. Images retrieved by the keyword expansion (red apple) are much fewer unrelated than those regained by the unique keyword (apple). The recovered top images are used as the working out examples of the reference class after robotically eliminating outliers. some reference classes (such as apple laptop and apple macbook) have related semantic meanings and their exercise sets are visually related. The redundant reference classes are removed in order to increase the proficiency of online image re-ranking. To better measure the similarity of semantic signatures, the semantic correlation between reference classes is estimated with a web-based kernel function. For each query keyword, its reference classes forms the basis of its semantic space. A multi-class classifier on visual and textual features is skilled from the exercise sets of its reference classes and deposited offline. Under a query keyword, the semantic signature of an image is extracted by computing the resemblances amongst the image and the reference classes of the query keyword using the trained multiclass classifier. If there are K types of visual/textual features, such as color, texture, and shape, one could combine them together to train a single classifier, which extracts one semantic signature for an image. A separate classifier for each type of feature can also be trained. Then, the K classifiers based on different types of features extract K semantic signatures, which are combined at the later stage of image matching. An image may be associated with multiple query key-words, which have different semantic spaces affording to the word- image index file. Hence, it may have different semantic signatures. The query keyword input by the user decides which semantic signature to choose. As an example shown in Fig. 2, an image is associated with three keywords apple, mac and computer. When using any of the three keywords as query, this image will be retrieved and re-ranked. However, under different query keywords, different semantic spaces are used. Therefore an image could have several semantic signatures obtained in different semantic spaces. They all need to be calculated and deposited offline.

At the online point, the search engine, affording to the query keyword, regains a pool of images. Meanwhile all the images in the pool are connected with the query keyword according to the word-image index file; they all have pre-computed semantic signatures in the same semantic space identified by the query keyword. Once the user chooses a query image, these semantic signatures are used to compute image similarities for re-ranking. The semantic correlation of reference classes is incorporated when computing the similarities and interaction is simple enough. Major web image search engines have adopted this strategy. Its diagram is shown in Fig. 1. Given a query keyword input by a user, a pool of images relevant to the query keyword is retrieved by the search engine according to a stored word-image index file. Usually the size of the returned image pool is fixed, e.g., containing 1000 images.

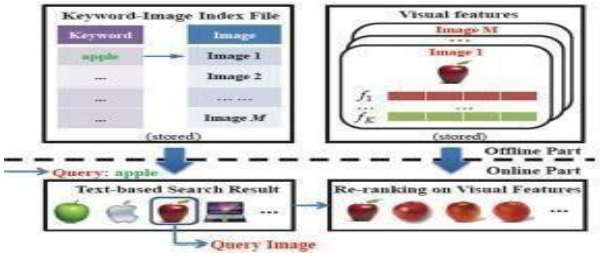


Fig. 1. The conventional framework for image re-rankinThe user is asked to select a query image from the pool.

This image reflects the user's search intention and

the remaining images in the pool are re-ranked based on their visual similarities with the query image. The word-image index file and visual features of images are pre-computed offline and stored. The main online computational cost is on comparing visual features. To achieve high efficiency, the visual feature vectors need to be short and their matching needs to be fast. Some popular visual features are in high dimensions and efficiency is not satisfactory if they are directly matched. In the current approaches, all the concepts/ attributes/ reference-classes are universally applied to all the images and they are manually defined. They are more suitable for offline databases with lower diversity (such as animal data- bases and face databases), since image classes in these databases can share similarities in a better way. A huge set of concepts or reference classes are required to model all the web images, which is impractical and ineffective for online image re- ranking. Intuitively, only a small subset of the concepts is relevant to a specific query.

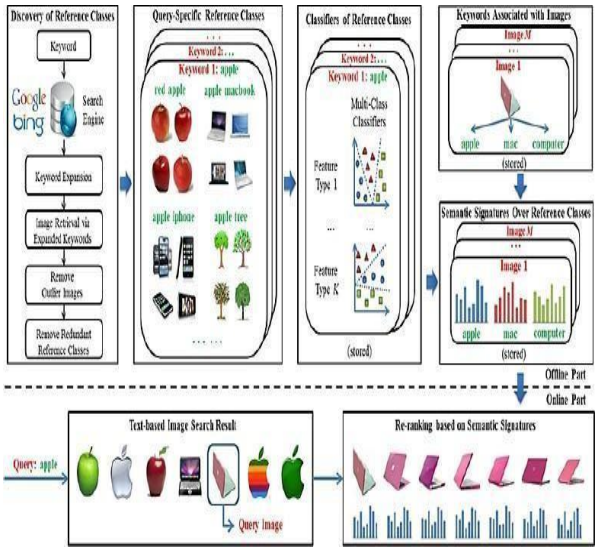


Fig.2.A new image re-ranking frame work

The conventional framework compares images based upon their visual features. The length of these visual features is much longer than that of the semantic signatures used in the new framework. Hence, the computational cost is higher. Compared with the conventional image re-ranking diagram in Fig. 1, the new approach is much more efficient at the online stage, because the main online computational cost is on comparing semantic signatures and the lengths of semantic signatures are much shorter than those of low-level visual features.

IV. SYSTEM DESIGN

Figure 3 demonstrates the framework of our proposed approach. In that user have to enter the query keyword into the search engine. Then search engine will return thousands of images based on text-based search. At that point our new framework will do re-rank the images by the following modules:

A. Keyword Expansion

Database keyword search (DB KWS) has received a lot of attention in the database research community. Although much of the research has been motivated by improving performance, recent research has also paid increased attention to its role in the database contents exploration or data mining. In this paper, we explore aspects related to DB KWS in two steps: First, we expand DB KWS by incorporating ontologies to better capture users' intention. Furthermore, we examine how KWS or ontology-enriched KWS can offer useful hints for better understanding of the data and in-depth analysis of the data contents, or data mining

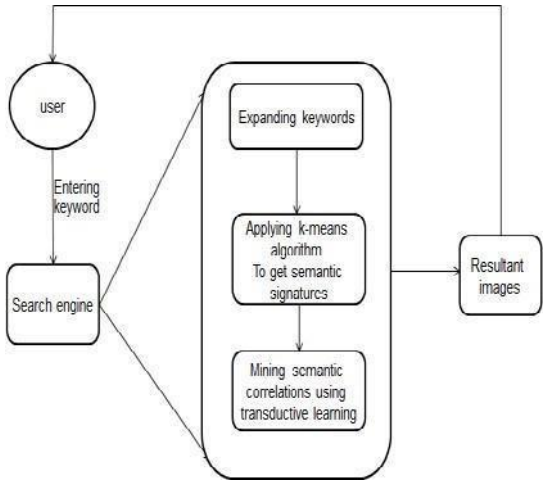


Fig. 3. System architecture

B. Semantic Signatures

Choosing keywords plays a main role in scheming semantic signatures; careful choice of keywords leads to a more accurate analysis, especially in English, which is sensitive to semantics. It is interesting to note that when words appear in different contexts they carry a different meaning. We have incorporated stemming within the framework and its effectiveness is demonstrated using a large corpus. We haveconducted experiments to demonstrate the sensitivity of semantic signatures to subtle content

differences between closely related documents. These experiments show that the newly developed

framework can identify subtle semantic differences substantially.

C. Semantic Correlations

Although multimedia objects such as images, audios and texts are of different modalities, there is a great amount of semantic correlations among them. In this paper, we propose a method of transductive learning to mine the semantic correlations among media objects of different modalities so that to achieve the cross-media retrieval. Cross-media retrieval is a new kind of searching technology by which the query examples and the returned results can be of different modalities, e.g., to query images by an example of audio. First, according to the media object features and their co-existence information, we construct a uniform cross-media correlation graph, in which media objects of different modalities are represented uniformly.

To perform the cross-media retrieval, a positive score is assigned to the query example; the score spreads along the graph and media objects of target modality or MMDs with the highest scores are returned. To boost the retrieval performance, we also propose different approaches of long-term and short-term relevance feedback to mine the information contained in the positive and negative examples.

V. CONCLUSION

In this paper, we have reviewed an Internet based image search approach. We have also discussed the conventional web-based image search techniques and pointed out their shortcomings. The reviewed image re-ranking framework overcomes the shortcomings of the previous methods and also significantly increases together the accuracy and efficiency of the re-ranking procedure. It captures users' intention using a query image. It learns query-specific semantic spaces to significantly improve the effectiveness and efficiency of online image re-ranking. The visual features of images are estimated into their related semantic spaces mechanically learned through keyword expansions offline. The extracted semantic signatures are shorter than the original visual features. In future work, image re-ranking can be further improved by incorporating other metadata and log data along with the textual and visual features for finding the keyword expansions used for defining the reference classes. The log data of user queries provides useful co-occurrence information of keywords for keyword expansion. Finally, in order to further improve the quality of re-ranked images, they should be re-ranked not only by content similarity but also by the visual quality of the images.

REFERENCES

[1] X. Tang, K. Liu, J. Cui, F. Wen, and X. Wang, (2012), Intent Search: Capturing User Intention for One-Click Internet Image Search, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 34, no. 7, pp. 1342-1353.

[2] J. Cui, F. Wen, and X. Tang, (2008), Real Time Google and Live Image Search Re-Ranking, Proc. 16th ACM Int'l Conf. Multimedia.

[3] Xiaogang Wang; Shi Qiu; Ke Liu; Xiaoou Tang, (2014), Web Image Re-Ranking UsingQuery-Specific Semantic Signatures, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.36, no.4, pp.810,823,doi: 10.1109/TPAMI.2013.214.

[4] X. Wang, K. Liu, and X. Tang, (2010), Query-Specific Visual Semantic Spaces for Web Image Re-Ranking, Proc. IEEE Conf. Computer Vision and Pattern Recognition(CVPR).

[5] Y. Jing and S. Baluja, (2008), Visual Rank: Applying Page Rank to Large-Scale Image Search, IEEE Trans. Pattern Analysis and Machine Intelligence,vol. 30, no. 11, pp. 1877-1890.

[6] J. Cai, Z. Zha, W. Zhou, and Q. Tian, (2012), Attribute-Assisted Reranking for Web Image Retrieval, Proc. 20th ACM Int'l Conf. Multimedia.

[7] N. Rasiwasia, P.J. Moreno, and N. Vasconcelos, (2007), Bridging the Gap: Query by Semantic Example, IEEE Trans. Multimedia, vol. 9, no. 5, pp. 923-938.

[8] C. Lampert, H. Nickisch, and S. Harmeling, (2009), Learning to Detect Unseen Object Classes by Between-Class Attribute Transfer, Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR).

[9] D. Parikh and K. Grauman, (2011), Relative Attributes, Proc. Int'l Conf. Computer Vision (ICCV).

[10] L. Torresani, M. Szummer, and A. Fitzgibbon, (2010), Efficient Object Category Recognition using Classemes, Proc. European Conf. Co

ICADET: conference proceedings: 2024 Advanced Development in Engineering And Technology

ISSN: 2454-9924

Cloud Based Data Recovery and Reconstruction System using Bi Methodology Erasure Code Implementation

P. Praveen Kumar #1, K. Madhan *1

Mailam Engineering College, Mailam #1, *1

pkumartin@gmail.com #1

Abstract - Reconstruction time has been minimized in erasure coded cloud storage. In previous work, as per the Traditional Reconstruction Techniques, Master node sends the request to the Worker node dedicated for the Reconstruction Process. This process encounters lots of Bottleneck Problems. In the proposed method, we are implementing Two Techniques namely, PUSH–Rep & PUSH–Sur. In PUSH–Rep Reconstruction occurs using Replacement Nodes. Rebuilt blocks are sequentially written to the disks of replacement nodes. PUSH–Sur allows each surviving node to rebuild a subset of failed data, so all the surviving nodes accomplish the reconstruction in parallel. In modified work, we are deploying this Application in Cloud. Data is encrypted, separated and stored in different Cloud. Replica is created for data backup. Top Hash Key is stored in Separate Cloud as well in the Local Backup. We implement PUSH–Rep using reconstruction from Cloud Backup and PUSH–Sur reconstruction from Local Backup.

Keywords – Erasure code, replacement, reconstruction, Parallel

1 Introduction

Traditional reconstruction techniques in storage clusters advocate the pull model, where a master node initiates reconstruction by sending requests to worker nodes

ISSN: 2454-9924

dedicated to the reconstruction process. The passive pull model inevitably encounters a transmission bottleneck problem that lies in rebuilding nodes. In this paper, we propose two PUSH-based reconstruction schemes

ISSN: 2454-9924

PUSH-Rep and PUSH-Sur—to improve reconstruction performance in a distributed storage cluster. At the heart of this study is the proactive PUSH technique that evenly distributes network and I/O loads among surviving nodes to shorten reconstruction times. The following three factors motivate us to propose the PUSH-based reconstruction technique for erasure-coded clustered storage. The high cost-effectiveness of erasure-coded storage, the severe impact of recovery time on reliability, and the deficiency of PULL-based reconstruction of input and output. 1. Erasure-coded storage clusters have increasingly become a cost-effective and fault-tolerant solution for archive storage data centers cloud storage and the like. Especially, Reed-Solomon (RS) codes are widely used in storage clusters to provide high data reliability. For example, Windows Azure Storage (WAS) adopts a variant of RS codes to implement a four-fault-tolerant cluster system. A detailed review on the RS-coded distributed storage is provided in Motivation 2. Ideally, erasure-coded storage clusters should protect against data loss caused by node failures, because high

ISSN: 2454-9924

reliability is an indispensable requirement for building large-scale storage systems. The mean-time-to data-loss or MTTDL of a r- fault-tolerant storage system is inversely proportional to the power of recovery time of a storage node. Therefore, it is extremely important to speed up the reconstruction process, which in turn can improve system reliability by shrinking vulnerability window size. The existing reconstruction schemes adopt a PULL-transmission mode, where a rebuilding node initiates the reconstruction by sending read requests to fetch/pull surviving blocks. Such a PULL mode not only raises the TCP In cast problem due to its synchronized many-to-one traffic pattern, but also yields poor reconstruction performance. When it comes to a reconstruction which relies on replacement nodes, the network traffic of replacement nodes contributes to an excessively long reconstruction time. The problem with the reconstruction among surviving nodes is thateach surviving node bears extra seek time owing to the non-contiguous disk access. This problem makes the low write bandwidth become a major reconstruction performance bottleneck. In this paper, we

ISSN: 2454-9924

introduce a PUSH-type transmission to speed up node-reconstruction performance. Our PUSH enables surviving nodes to accomplish reconstruction tasks in a pipelining manner. Each surviving node combines its local block with an intermediate block from another surviving node to partially generate an intermediate block forwarded to a subsequent node. Thus, PUSH can speed up the reconstruction process by maximizing the utilization of both network and I/O bandwidth of all the surviving nodes.

2 Related Work

In the previous work, as per the Traditional Reconstruction Techniques, Master node sends the request to the Worker node dedicated for the Reconstruction Process. This process encounters lots of Bottleneck Problems. Here it provides some drawbacks are, Waiting time is increased, Congestion occurring, Unreliable, Less data transmission rate Less effective

2.1 Proposed Work

In the proposed work, we are implementing Two Techniques namely, PUSH–Rep & PUSH–Sur. In PUSH–Rep Reconstruction

ISSN: 2454-9924

occurs using Replacement Nodes. Rebuilt blocks are sequentially written to the disks of replacement nodes. PUSH–Sur allows each surviving node to rebuild a subset of failed data, so all the surviving nodes accomplish the reconstruction in parallel. In the modified work, we are deploying this Application in Cloud. Data is encrypted, separated and stored in different Cloud. Replica is created for data backup. Top Hash Key is stored in Separate Cloud as well in the Local Backup. We implement PUSH– Rep using reconstruction from Cloud Backup and PUSH–Sur reconstruction from Local Backup.

3 PULL-Based Reconstruction Scheme Let us consider two existing reconstructiontechniques that rely on the pull mode, wherea rebuilding node first issues read requests tosurviving nodes and then reconstructs afailed block using the requested blocks. ThePULL-based reconstruction can beenvisioned as a master-worker computingmodel, in which a real-

ISSN: 2454-9924

world erasure-coded storage clusters: i) a designated master (e.g., a replacement node) fetches k surviving blocks and reconstructs a failed block, and ii) each surviving node plays the role of a master (i.e., acting as a rebuilding node) and all surviving nodes perform as workers, where write I/O s of rebuilt blocks are spread out over all the surviving nodes. From the angle of message communication, this ‗Master Worker' pattern belongs to the category of PULL-type transmission. Throughout this paper, we refer to the reconstruction scheme using replacement nodes as PULL Rep; we term the solution of distributing reconstruction load among surviving nodes as PULL-Sur. In the case of PULL-Rep, all reconstruction reads are sequential requests that minimize disk seeking times; rebuilt blocks are sequentially written to disks of replacement nodes. Fig. 2a shows that k surviving blocks should be delivered to a replacement node (e.g., RN), which becomes a network bottleneck that slows down the entire reconstruction process. Furthermore, such a many- to-one (‗M: 1') communication pattern may cause the severe In cast problem

4 Literature Review

[4] Describes about, Digital archives are growing rapidly, necessitating stronger reliability measures than RAID to avoid data loss from device failure. Mirroring, a popular solution, is too expensive over time. We present a compromise solution that uses multi-level redundancy coding to reduce the probability of data loss from multiple simultaneous device failures. This approach handles small-scale failures of one or two devices efficiently while still allowing the system to survive rare-event, larger-scale failures of four or more devices. In our approach, each disk is split into a set of fixed size diskless which are used to construct reliability stripes. To protect against rare event failures, reliability stripes are grouped into larger –¨user-groups,‖ each of which has a corresponding –¨user-parity;‖ ¨user-parity is only used to recover data when disk failures overwhelm the redundancy in a single reliability stripe. ¨User-parity can be stored on a variety of devices such as NV-RAM and always-on disks to offset write bottlenecks while still keeping the number of active devices low.

ISSN: 2454-9924

Our calculations of failure probabilities

ISSN: 2454-9924

found that the addition of ¨user-groups allowed the system to absorb many more disk failures without data loss. Through discrete event simulation, we found that adding ¨user-groups only negatively impacts performance when these groups need to be used for a rebuild. Since rebuilds using ¨user-parity occur very rarely, they minimally impact system performance over time. Finally, we showed that robustness against rare events can be achieved for fewer than 5% of total system cost. [16] Describes about, In this paper we describe Cumulus, a system for efficiently implementing file system backups over the Internet. Cumulus is specifically designed under a thin cloud assumption—that the remote datacenter storing the backups does not provide any special backup services, but Only provides a least-common-denominator storage interface (i.e., get and put of complete files). Cumulus aggregates data from small files for remote storage, and uses LFS-inspired segment cleaning to maintain storage efficiency. Cumulus also efficiently represents incremental changes, including edits to large files. While Cumulus can use virtually any storage service, we show that

ISSN: 2454-9924

its efficiency is comparable to integrated approaches. [22] Describes about, In spite ofthe central role of key derivation functions (KDF) in applied cryptography, there hasbeen little formal work addressing the design and analysis of general multi-purpose KDFs. In practice, most KDFs (including those widely standardized) follow ad-hoc approaches that treat cryptographic hash functions as perfectly random functions. In this paper we close some gaps between theory and practice by contributing to the study and engineering of KDFs in several ways. We provide detailed rationale for the design of KDFs based on the extract- then- expand approach; we present the first general and rigorous dentition of KDFs and their security that we base on the notion of computational extractors; we specify a concrete fully practical KDF based on the HMAC construction; and we provide an analysis of this construction based on the extraction and pseudorandom properties of HMAC. The resultant KDF design can support a large variety of KDF applications under suitable assumptions on the underlying hash function; particular attention and effort is devoted. [23]

ISSN: 2454-9924

Describes about, The increasing popularity of cloud storage is leading organizations to consider moving data out of their own data centers and into the cloud. However, success for cloud storage providers can present a significant risk to customers; namely, it becomes very expensive to switch storage providers. In this paper, we make a case for applying RAID-like techniques used by disks and file systems, but at the cloud storage level. We argue that striping user data across multiple providers can allow customers to avoid vendor lock-in, reduce the cost of switching providers, and better tolerate provider outages or failures. We introduce RACS, a proxy that transparently spreads the storage load over many providers. We evaluate a prototype of our system and estimate the costs incurred and benefits reaped. Finally, we use trace-driven simulations to demonstrate how RACS can reduce the cost of switching storage vendors for a large organization such as the Internet Archive by seven-fold or more by varying erasure-coding parameters. [7] Describes about, Latent sector errors (LSEs) refer to the situation where particular sectors on a drive

ISSN: 2454-9924

become inaccessible. LSEs are a critical factor in data reliability, since a single LSE can lead to data loss when encountered during RAID reconstruction after a disk failure. LSEs happen at a significant rate in the field [1], and are expected to grow more frequent with new drive technologies and increasing drive capacities. While two approaches, data scrubbing and intra-disk redundancy, have been proposed to reduce data loss due to LSEs, none of these approaches has been evaluated on real field data. This paper makes two contributions. We provide an extended statistical analysis of latent sector errors in the field, specifically from the view point of how to protect against LSEs. In addition to providing interesting insights into LSEs, we hope the results (including parameters for models we fit to the data) will help researchers and practitioners without access to data in driving their simulations or analysis of LSEs. Our second contribution is an evaluation of five different scrubbing policies and five different intra-disk redundancy schemes and their potential in protecting against LSEs. Our study includes schemes and policies that have been

ISSN: 2454-9924

suggested before, but have never been evaluated on field data, as well as new policies that we propose based on our analysis of LSEs in the field.

6 Methodologies

6.1 Owner in cloud

User is the person is going to see or download the data from the Cloud server. To access the data from the Cloud server, the users have to be registered with the cloud server. So that the user have to register their details like username, password and a set of random numbers. This is information will stored in the database for the future authentication. Data Owner: Data Owner is the Person who is going to upload the data in the Cloud Server. To upload the data into the Cloud server, the Data Owner have be registered in the Cloud Server. Once the Data Owner registered in cloud server, the space will be allotted to the Data Owner.

6.2 Cloud Server of main

Cloud Server is the area where the user going to request the data and also the data owner will upload their data. Once the user send the request regarding the data they want, the request will first send to the Cloud Server and the Cloud Server will forward your request to the data owner. The data Owner will send the data the data the user via Cloud Server. The Cloud Server will also maintain the Data owner and Users information in their Database for future purpose.

6.3 Partition of data and encryption

In this module, once the data was uploaded into the cloud server, the Cloud server will split the data into many parts and store all the data in the separate data servers. In techniques wasn‘t used in proposed system so that there might be a chance of hacking the entire data. Avoid the hacking process, we splitting the data and store those data in corresponding data server. We‘re also encrypting the data segments before storing into the data server.

6.4 Key production Server

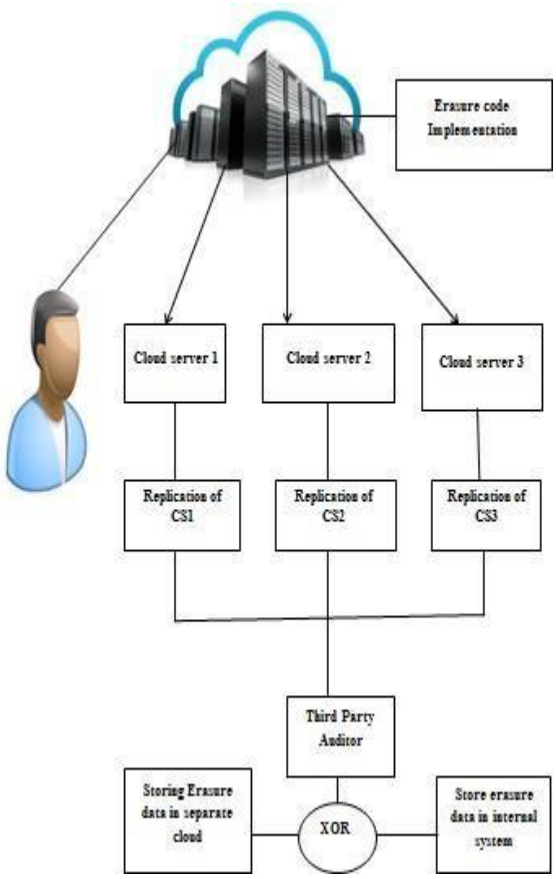
The encryption keys are stored in appropriate key servers. So that we can increase the security of the cloud network. If

ISSN: 2454-9924

the user wants retrieve the data, they‘ve to provide all the key that are stored in the appropriate key servers.

7 System Design

The architecture mainly based on the pull based technique, The PULL-based reconstruction can be envisioned as a master-worker computing model, in which a master triggers a reconstruction procedure by sending a set of read requests, and then waits for the requests to be completed by workers.



There are two classical reconstruction approaches in real-world erasure-coded storage clusters: i) a designated master (e.g., a replacement node) fetches k surviving blocks and reconstructs a failed block; and ii) each surviving node plays the role of a master (i. e., acting as a rebuilding node) and all surviving nodes perform as workers, where write I/O s of rebuilt blocks are spread out over all the surviving nodes.

ISSN: 2454-9924

From the angle of message communication,

ISSN: 2454-9924

this _Master Worker‘ pattern belongs to the category of PULL-type transmission. Throughout this paper, we refer to the reconstruction scheme using replacement nodes as PULL Rep. Also it depends on the The goal of the PUSH technique for node reconstruction is two-fold. First, PUSH aims to alleviate the reconstruction performance bottleneck caused by a replacement node‘s network bandwidth in PULL-Rep. Second, PUSH also aims to mitigate extra seeking times induced by the non-sequential disk accesses in PULL-Sur. In comparison to surviving nodes that passively respond to reconstruction reads in PULL, the surviving nodes in PUSH proactively participate in the entire reconstruction process.

8 Conclusion

From this, Cloud Based Data Recovery and Reconstruction System using Bi Methodology Erasure Code Implementation have been implemented. Nowadays a grand challenge for storage clusters is efficiently migrating data replicas to create an erasure-coded archive. To take this challenge, we are going to integrate the PUSH-type

ISSN: 2454-9924

transmission into the archival migration in erasure-coded storage clusters. Moreover, since PUSH-based reconstruction schemes are sensitive to slow nodes, we plan to extend the PUSH-based reconstruction schemes for heterogeneous erasure-coded storage clusters by taking into account both load and heterogeneity of surviving nodes. To address these issues, we proposed the PUSH approach, in which a PUSH-type transmission is incorporated into node reconstruction. We developed two PUSH- based reconstruction schemes (i.e., PUSH Rep and PUSH-Sur). Compared to the PULL-based counterparts where surviving blocks are transferred in a synchronized _M:1‘ traffic pattern, our PUSH-based reconstruction solutions support the _1:1‘ pattern, which naturally solves the In cast problem. We built performance models to investigate the reconstruction times of our PUSH-based schemes applied in large-scale storage clusters. We extensively evaluated the four schemes on a real-world cluster.

9 References

[1] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran,

ISSN: 2454-9924

–Network coding for distributed storage systems,‖ IEEE Trans. Inform. Theory, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.

[2] A. Kermarrec, N. Le Scouarnec, and G. Straub, –Repairing multiple failures with coordinated and adaptive regenerating codes,‖ in Proc. Int. Symp. Netw. Coding, 2011, pp. 1–6.

[3] A. Phanishayee, E. Krevat, V. Vasudevan, D. Andersen, G. Ganger, G. Gibson, and S. Seshan, –Measurement and analysis of TCP throughput collapse in cluster-based storage systems,‖ in Proc. 6th USENIX Conf. File Storage Technol., 2008, p. 12.

[4] Avani Wildani , Protecting Against Rare Event Failures in Archival Systems, April 2009

[5] B. Cassidy, J. Hafner, Space efficient matrix methods for lost data reconstruction in erasure codes,‖ IBM Res., Armonk, NY, USA, Tech. Rep. RJ10415, 2007.

[6] B. Calder et al., –Windows azure storage: A highly available cloud storage service with strong consistency,‖ in Proc.

23rd ACM Symp. Operating Syst. Principles, 2011, pp. 143–157.

[7] Bianca Schroeder, Understanding latent sector errors and how to protect against them, coding have a role to play in my data center?‖ Microsoft research MSR-TR-2010, vol. 52, 2010.

[8] B. Welch, M. Unangst, Z. Abbasi, G. Gibson, B. Mueller, J. Small, J. Zelenka, and
B. Zhou, ―Scalable performance of the panasas parallel file system,‖ in Proc. 6th USENIX Conf. File Storage Technol.,vol.2,2008,pp.1–2.

[9] C. Huang, H. Simitci, Y. Xu, A. Ogus, B. Calder, P. Gopalan, J. Li, and S. Yekhanin, ―Erasure coding in windows azure storage,‖ in Proc. USENIX Annu. Tech. Conf., 2012, p. 2.

[10] C. Dubnicki, L. Gryz, L. Heldt, M. Kaczmarczyk, W. Kilian, P. Strzelczak, J. Szczepkowski, C. Ungureanu, and M. Welnicki, ―Hydrastor: A scalable secondary storage,‖ in Proc. 7th Conf. File Storage Technol., 2009, pp. 197–210.

ISSN: 2454-9924

[11] I. Reed and G. Solomon, ―Polynomial codes over certain finite fields,‖ J. Soc. Ind. Appl. Math., vol. 8, no. 2, pp. 300–304, 1960.

[12] J. Plank et al., ―A tutorial on reed-solomon coding for fault-tolerance in raid-like systems,‖ Softw. Practice Experience, vol. 27, no. 9, pp. 995–1012, 1997.

[13] K. Rao, J. Hafner, and R. Golding, ―Reliability for networked storage nodes,‖ IEEE Trans. Dependable Secure Comput., vol. 8, no. 3, pp. 404–418, May 2011.

[14] L. Xiang, Y. Xu, J. Lui, and Q. Chang, ―Optimal recovery of single disk failure in rdp code storage systems,‖ ACM SIGMETRICS Per- form. Eval. Rev., vol. 38, no. 1, pp. 119–130, 2010.

[15] M. Aguilera, R. Janakiraman, and L. Xu, ―Using erasure codes efficiently for storage in a distributed system,‖ in Proc. Int. Conf. Dependable Syst. Netw., 2005, pp. 336–345.

[16] Michael Vrable,Cumulus: File system Backup to the Cloud,

[17] M. Holland, G. Gibson, and D. Siewiorek, ―Fast, on-line failure recovery in redundant disk arrays,‖ in Proc. 23rd Int. Symp. FaultTolerant Comput., 1993, pp. 422–431.

[18] M. Holland, G. Gibson, and D. Siewiorek, ―Architectures and algorithms for on-line failure recovery in redundant disk arrays,‖ Distrib. Parallel Databases, vol. 2, pp. 295–335, 1994.

[19] M. Manasse, C. Thekkath, and A. Silverberg, ―A reed-solomon code for disk storage, and efficient recovery computations for erasure-coded disk storage,‖ Proc. Inf., pp. 1–11, 2009.

[20] O. Khan, R. Burns, J. Plank, W. Pierce, and C. Huang, ―Rethinking erasure codes for cloud file systems: Minimizing I/O for recovery and degraded reads,‖ in Proc. 10th USENIX Conf. File Storage Technol., 2012, pp. 251–264.

ISSN: 2454-9924

[21] Q. Xin, E. Miller, T. Schwarz, D. Long, S. Brandt, and W. Litwin, ―Reliability mechanisms for very large storage systems,‖ in Proc. 20th IEEE/11th NASA Goddard Conf. Mass Storage Syst. Technol., 2003, pp. 146–156.

[22] T.J. Watson Research Center,Cryptographic Extraction and Key Derivation: The HKDF Scheme, 2010

[22] Q. Xin, E. Miller, and S. Schwarz, ―Evaluation of distributed recovery in large-scale storage systems,‖ in Proc. 13th IEEE Int. Symp. High Performance Distrib. Comput., 2004, pp. 172–181.

[23] RACS: A Case for Cloud Storage Diversity

ISSN: 2454-9924

CLOUD COMPUTING SERVICE FOR DYNAMIC DATA STORAGE SYSTEM

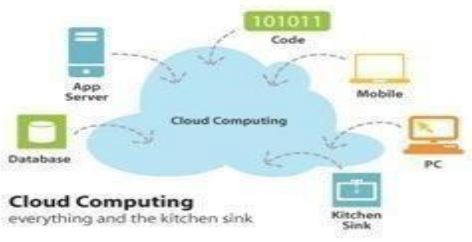
K. Rajkumar*, Dr. G. Shanmugasundaram #
Sri ManakulaVinayagar Engineering College * #
rajpjp2011@gmail.com *

Abstract:

The offered idea of virtualization provides improved system utilization via virtual infrastructure and promotes resource sharing across an party Cloud computing is a in recent times evolved computing terms or simile based on utility and using up of computing resources. Cloud computing involves deploy groups of remote servers and software networks that allow centralized data storage and online networks be access to computer services or resources in support of case, a cloud computer capability that serves European users during European business hours with a specific application.

INTRODUCTION:

Cloud Computing is the internet-based storage space for files, applications, and infrastructure. One could declare cloud computing has been around for many in years, but now a company may buy or rent space for their daily operations. The cost savings in implementing a cloud system is significant and the pricing for use of cloud computing can easily be scaled up or down as resolute by requirement.



Service models

ISSN: 2454-9924

Cloud computing providers propose theirservices according to three fundamental models: infrastructure as a service and platform as a service , and software as a service providers in cloud network where place IaaS is the most fundamental and each higher model abstract from the details of the lower models were officially included by ITU (International Telecommunication Union) as part of the basic cloud computing models, standard service categories of a telecommunication-centric cloud Bionetwork.

Infrastructure as a service (IaaS)

In the most primitive cloud-service model, providers of IaaS offer computers - physical or (more often) virtual machines - and new

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN: 2454-9924

resources. To deploy their applications, cloud users install operating-system similes and their application software on the cloud infrastructure. here model, the cloud user patch and maintain the operating systems and the application software. Cloud providers usually bill IaaS services on a service computing basis: cost reflect the amount of resources to be paid and consumed.

Platform as a service (PaaS)

(PaaS) is a kind of cloud computing services that provides a platform allow customers to develop, run and manage Web applications with no the simplicity of building and maintaining the infrastructure in general associated with developing and debut an app. PaaS can be delivered in two ways: as a public cloud service from a supplier, where the customer controls software employment and configuration setting, and the provider cloud networks, servers, storage and new services to host the clients application; or as software installed in private data centers or public infrastructure as a service and administerd by inside IT department

Software as a service (SaaS)

SaaS) has brought a huge difference in the customs in which business is done today. As we identify, Cloud Computing is a service through which you can reward shared resources, software and information on your computer or other devices via the Internet. This resources that you can access the data rations you want any time, price your claims on the go, save time on tiresome reporting, claims agreement and much more.

Network as a service (NaaS)

Network as a Service (NaaS) is sometimes listed as a divide Cloud provider services along with Infrastructure as a Service (IaaS), Platform as a Service provides them the same traditional features with a faster speed, reducing the cost and increasing collaboration Platform as a Service (PaaS), and Software as a Service (SaaS). This factor out networking, firewalls connected security, etc.

Deployment models:

Private cloud

Private cloud is cloud infrastructure operate exclusively for a single organization, whether managed internally or by a third-party, and hosted either on the inside or externally. Undertaking a private cloud project requires a major level and degree of commitment to virtualize the business environment, and requires the group to revaluate decisions about existing resources., it can improve business, but every step in the project raise security issue that must be addressed to prevent serious vulnerabil ities**.**

Community cloud

Community cloud shares infrastructures ure between several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted internally or externally. The costs are spread over fewer users hana public cloud (but more

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN: 2454-9924

than a private cloud), so only some of the cost savings potential of cloud computing are realized.

Hybrid cloud:

Hybrid cloud is a work of two or more clouds (private, community or public) that remain diverse entities but are bound together, offering the payback of multiple deployment models. Hybrid cloud can also mean the ability to connect collocation, manage and/or committed services with cloud capital. For example, an society may store sensitive client data in ho use on a private cloud application, but communicate that application to a business aptitude applicatio

The Intercloud

The Intercloud is an organized global "cloud of clouds" and an extension of the Internet "network of networks" on which it is based. The focal point is on direct interoperability between public cloud service providers, more so than between providers and customers cloud

Engineering

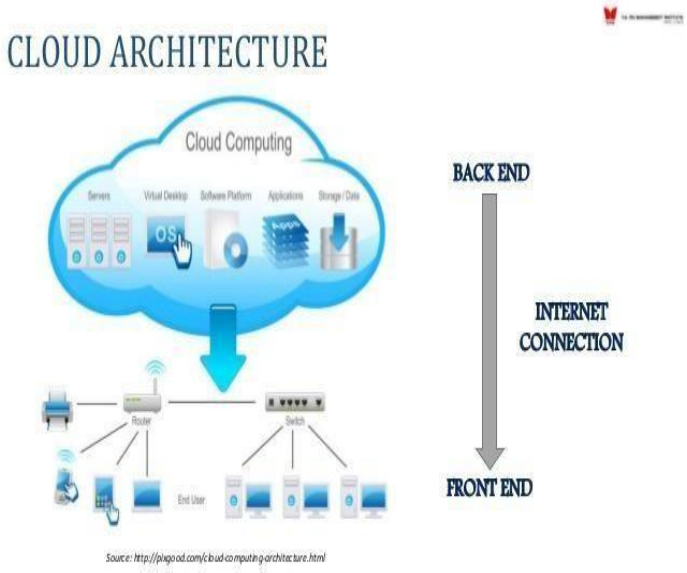
Cloud engineering is the application of engineering discipline to cloud computing. It brings a systematic approach to the high-level concern of commercialization, standardisation, and governance in conceive, developing, systems, cost, software, web, performance, information, security, platform, risk, and quality cloud engineering.

n provided on a community cloud as a software service.[70] This example of hybrid cloud extend the potential of the project to deliver a specific business service through the calculation of externally available public cloud services

Architecture:

Cloud architecture, of the software systems complex in the delivery of cloud computing, usually involves multiple *cloud components* communicating with each other over a free union mechanism such as a messaging queue. Elastic provision implies intellect in the use of tight or free union as functional to mechanisms such as these and others.

operating and maintaining cloud computing systems. It is a multi corrective method about contributions from diverse areas such as



ISSN: 2454-9924

Reference:

[1]Hassan , Qusay (2011). "Demystifying Cloud Computing". The Journal of Defense Software Engineering (CrossTalk) 2011 (Jan/Feb): 16–21. Retrieved11 December 2014.

[2] "The NIST Definition of Cloud Computing". National Institute of Standards and Technology. Retrieved 24 July 2011.

[3] "What is Cloud Computing?". Amazon Web Services. 2013-03-19. Retrieved 2013-03-20.

[4] "Baburajan, Rajani, "The Rising Cloud Storage Market Opportunity Strengthens Vendors," infoTECH, August 24, 2011". It.tmcnet.com. 2011-08-24.

Economist. 2009-10-15. Retrieved 2009-11-03. [8]"Gartner Says Cloud Computing Will Be As Influential As E-business". Gartner. Retrieved 2010-08-22.

[9]Gruman, Galen (2008-04-07). "What cloud computing really means". InfoWorld. Retrieved 2009-06-02. [10]"The economy is flat so why are financials Cloud vendors growing at more than 90 percent per annum?". FSN. March 5, 2013.

[11] Hongji Yang, Xiaodong (2012). "9". Software reuse in the emerging cloud computing era. Hershey, PA: Information Science

[12] "A network 70 is shown schematically as a cloud", US Patent 5,485,455, column 17, line 22, filed Jan 28, 1994

Figure 1, "the cloud indicated at 49 in Fig. 1.", US Patent [13]5,790,548, column 5 line 56–57, filed April 18, 1996 Antonio Regalado (31 October 2011). "Who Coined 'Cloud Computing'?".Technology Review (MIT). Retrieved 31 July 2013.

ISSN: 2454-9924

Cloud services using supporting reputation based trust management

E. Prathipa #1, N. Geetha *1

Mailam Engineering College, Mailam #1,*1

Prathielumalai23@gmail.com #1

Abstract - In cloud computing growth, the management of trust element is most challenging issue. Cloud computing has produce high challenges in security and privacy by the changing of environments. Trust is one of the most concerned obstacles for the adoption and growth of cloud computing. Although several solutions have been proposed recently in managing trust feedbacks in cloud environments, how to determine the credibility of trust feedbacks is mostly neglected. In this project the system proposed a Cloud Armor, a reputation-based trust management framework that provides a set of functionalities to deliver Trust as a Service (TaaS). "Trust as a Service" (TaaS) framework to improve ways on trust management in cloud environments. The approaches have been validated by the prototype system and experimental results.

Keywords – Cloud computing, Trust, Obstacles, reputation, feedbacks

1 Introduction

The highly dynamic, distributed, and nontransparent nature of cloud services make the trust management in cloud environments a significant challenge. According to researchers at Berkeley, trust and security is ranked one of the top 10 obstacles for the adoption of cloud computing. Indeed, Service-Level Agreements (SLAs) alone are inadequate to establish trust between cloud consumers and providers because of its unclear and inconsistent clauses. Consumers' feedback is a good source to assess the overall trustworthiness of cloud services. Several researchers have recognized the significance of trust management and proposed solutions to assess and manage trust based on

feedbacks collected from participants. In reality, it is not unusual that a cloud service experiences malicious behaviors (e.g., collusion or Sybil attacks) from its users. This paper focuses on improving trust management in cloud environments by proposing novel ways to ensure the credibility of trust feedbacks. In particular we distinguish the following key issues of the trust management in cloud environments: Consumers' Privacy. The adoption of cloud computing raise privacy concerns .Consumers can have dynamic interactions with cloud providers, which may involve sensitive information. There are several cases of privacy breaches such as leaks of sensitive information (e.g., date of birth and address) or behavioral information (e.g., with whom the consumer interacted, the kind of cloud services the consumer showed interest, etc.). Undoubtedly, services which involve consumers' data (e.g., interaction histories) should preserve their privacy. Cloud Services Protection. It is not unusual that a cloud service experiences attacks from its users. Attackers can disadvantage a cloud service by giving multiple misleading feedbacks (i.e., collusion attacks) or by creating several accounts (i.e., Sybil attacks). Indeed, the detection of such malicious behaviors poses several challenges. Firstly, new users join the cloud environment and old users leave around the clock. This consumer dynamism makes the detection of malicious behaviors (e.g., feedback collusion) a significant challenge. Secondly, users may have multiple accounts for a particular cloud service, which makes it difficult to detect Sybil attacks. Finally, it is difficult to predict when malicious behaviors occur (i.e., strategic VS. occasional behaviors). Trust Management Service's Availability. A trust management service (TMS) provides an interface between users and cloud services for effective trust management. However, guaranteeing the availability of TMS is a difficult problem due to the unpredictable number of users and the highly dynamic nature of the cloud environment. Approaches that require understanding of users' interests and capabilities through similarity measurements or operational availability measurements (i.e., uptime to the total time) are inappropriate in cloud environments. TMS should be adaptive and

highly scalable to be functional in cloud environments.

2 Related Work

Trust is one of the most concerned obstacles for the adoption and growth of cloud computing. Although several solutions have been proposed recently in managing trust feedbacks in cloud environments, how to determine the credibility of trust feedbacks is mostly neglected. In this project the system proposed a Cloud Armor, a reputation-based trust management framework that provides a set of functionalities to deliver Trust as a Service (TaaS). "Trust as a Service" (TaaS) framework to improve ways on trust management in cloud environments. The approaches have been validated by the prototype system and experimental results. Here, it provides some drawbacks are, It is not unusual that a cloud service experiences malicious behaviors from its users, It is not sure whether they can trust the cloud providers, It not convincing enough for the consumers, SLAs are not consistent among the cloud providers even though they offer services with similar functionality,

Customers are not sure whether they can identify a trustworthy cloud provider only based on its SLA. In this project the system proposed a Cloud Armor, a reputation-based trust management framework that provides a set of functionalities to deliver Trust as a Service (TaaS) ."Trust as a Service" (TaaS) framework to improve ways on trust management in cloud environments. In particular, the system introduce an adaptive credibility model that distinguishes between credible trust feedbacks and malicious feedbacks by considering cloud service consumers' capability and majority consensus of their feedbacks. The approaches have been validated by the prototype system and experimental results. The system proposes a framework using the Service Oriented Architecture (SOA) to deliver trust as a service. Here it includes some benefits are, It not only preserves the consumers' privacy, but also enables the TMS to prove the credibility of a particular consumer's feedback, It also has the ability to detect strategic and occasional behaviors of collusion attacks, Load balancing techniques are exploited to share the workload, thereby always maintaining a

desired availability level, This metric exploits particle filtering techniques to precisely predict the availability of each node, Cloud Armor exploits techniques to identify credible feedbacks from malicious ones.

3 Literature Review

[13] describe about, In this paper we assess how security, trust and privacy issues occur in the context of cloud computing and discuss ways in which they may be addressed It has the advantage of reducing cost by sharing computing and storage resources, combined with an on-demand provisioning mechanism relying on a pay-per-use business model. This makes compliance with regulations related to data handling difficult to fulfill. [5] Describe about, We begin this paper with a survey of existing mechanisms for establishing trust, and comment on their limitations. We then address those limitations by proposing more rigorous mechanisms based on evidence, attribute certification, and validation, and conclude by suggesting a framework for integrating various trust mechanisms together to reveal chains of trust in the cloud. This system presents an integrated view of the trust mechanisms for cloud computing, and analyzes the trust chains connecting cloud entities. Some cloud clients cannot make decisions about employing a cloud service based solely on informal trust mechanisms. [7] describe about, The authors suggest using a trust-overlay network over multiple data centers to implement a reputation system for establishing trust between service providers and data owners. Data coloring and software watermarking techniques protect shared data objects and massively distributed software modules. These techniques safeguard multi-way authentications, enable single sign-on in the cloud, and tighten access control for sensitive data in both public and private clouds. Once users move data into the cloud, they can't easily extract their data and programs from one cloud serverto run on another. This leads to a data lock-in problem. [12] Describe about, the descriptions in SLAs are not consistent among the cloud providers even though the other services with similar functionality. Therefore, customers are not sure whether they can identify a trustworthy cloud

provider only based on its SLA. This system provides means to identify the trustworthy cloud providers in terms of different attributes assessed by multiple sources and roots of trust information; they are not sure whether they can trust the cloud providers. [9] In this paper, we tackle these problems by exploiting particle filtering-based techniques. In particular, we developed algorithms to accurately predict the availability of Web services and dynamically maintain a subset of Web services with higher availability ready to join service compositions. Web services can be always selected from this smaller space, thereby ensuring good performance in service compositions. Unfortunately, how to provide real-time availability information of Web services is largely overlooked.

4 Methodologies

4.1 Detection of service

This layer consists of different users who use cloud services. For example, a new startup that has limited funding can consume cloud services. Interactions for this layer include: i) service discovery where users are able to discover new cloud services and other services through the Internet, ii) trust and service interactions where users are able to give their feedback or retrieve the trust results of a particular cloud service, and iii) registration where users establish their identity through registering their credentials in IdM before using TMS.

4.2 Trust Communication

In a typical interaction of the reputation-based TMS, a user either gives feedback regarding the trustworthiness of a particular cloud service or requests the trust assessment of the service 1. From users' feedback, the trust behavior of a cloud service is actually a collection of invocation history records, represented by a tuple H= (C, S, F, T f), where C is the user's primary identity, S is the cloud service's identity, and F is a set of Quality of Service (QOS) feedbacks (i.e., the feedback represent several QOS parameters including availability, security, response time, accessibility, price).

4.3 IDM Registration

The system proposes to use the Identity Management Service (IdM) helping TMS in

measuring the credibility of a consumer's feedback. However, processing the IdM information can breach the privacy of users. One way to preserve privacy is to use cryptographic encryption techniques. However, there is no efficient way to process encrypted data. Another way is to use anonymization techniques to process the IDM information without breaching the privacy of users. Clearly, there is a trade-off between high anonymity and utility.

4.4 Service announcement and Communication

This layer consists of different cloud service providers who offer one or several cloud services, i.e., IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Soft-ware as a Service), publicly on the Web (more details about cloud services models and designs can be found). These cloud services are accessible through Web portals and indexed on Web search engines such as Google, Yahoo, and Baidu. Interactions for this layer are considered as cloud service interaction with users and TMS.

5 System Design

5.1 The Cloud Service Provider Layer

This layer consists of different cloud service providers who offer one or several cloud services, i.e., IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Software as a Service), publicly on the Web (more details about cloud services models and designs). These cloud services are accessible through Web portals and indexed on Web search engines such as Google, Yahoo, and Baidu. Interactions for this layer are considered as cloud service interaction with users and TMS, and cloud services advertisements where providers are able to advertise their services on the Web.

5.2 The Trust Management Service Layer

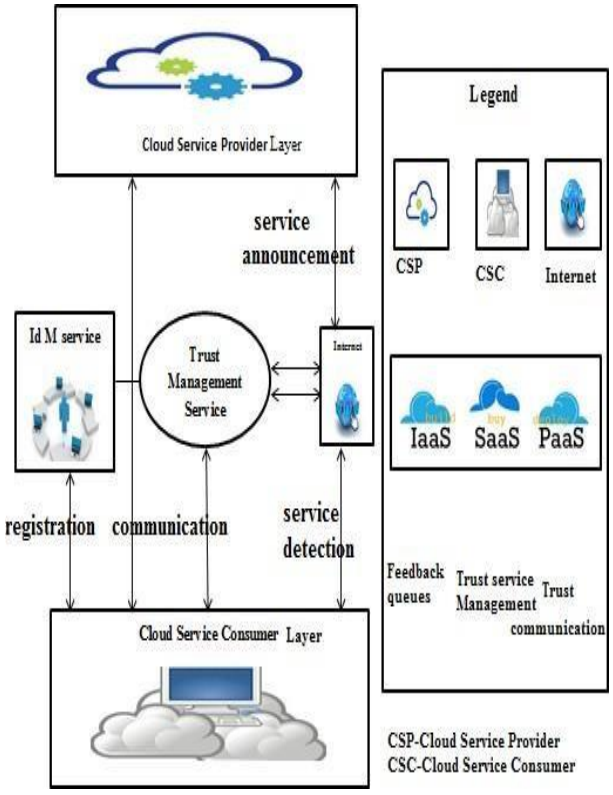
This layer consists of several distributed TMS nodes which are hosted in multiple cloud environments in different geographical areas. These TMS nodes expose interfaces so that users can give their feedback or inquire the trust results in a decentralized way. Interactions for this layer include:

i) cloud service interaction with cloud service providers, ii) service advertisement to advertise the trust as a service to users through the Internet, iii) cloud service discovery through the Internet to allow users to assess the trust of new cloud services, and iv) Zero-Knowledge Credibility Proof Protocol (ZKC2P) interactions enabling TMS to customers feedback.

5.3 The Cloud Service Consumer Layer

Finally, this layer consists of different users who use cloud services. For example, a new startup that has limited funding can consume cloud services (e.g., hosting their services in Amazon S3). Interactions for this layer include: i) service discovery where users are able to discover new cloud services and other services through the Internet, ii) trust and service interactions where users are able to give their feedback or retrieve the trust results of a particularcloud service, and iii) registration where users establish their identity through registering their credentials in IdM before using TMS. Our framework also exploits a Web crawling approach for automatic cloud services discovery, where cloud services are automatically discovered on the Internet and stored in a cloud services repository. Moreover, our framework contains an Identity Management Service, which is responsible for the registration where users register their credentials before using TMS and proving the credibility of a particular consumer's feedback through ZKC2P.

A service provider that includes customer storage or software services available through a private (private cloud) or public network (cloud).Usually, it means the storage and software is available for process through the Internet.



6 Conclusions

From this Cloud Armor Supporting Reputation-based Trust Management for Cloud Services has been implemented. In cloud computing growth, the management of trust element is most challenging issue. Cloud computing has produce high challenges in security and privacy by the changing of environments. Trust is one of the most concerned obstacles for the adoption and growth of cloud computing.

Although several solutions have been proposed recently in managing trust feedbacks in cloud environments, how to determine the credibility of trust feedbacks is mostly neglected. Additionally in future, we also enhance the performance of cloud as well as the security.

7 References

References

[1] A. Birolini, Reliability Engineering: Theory and Practice. Springer2010.

[2] C. Dellarocas, “The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms,” Management Science, vol. 49, no. 10, pp. 1407–1424, 2003.

[3] E. Bertino, F. Paci, R. Ferrini, and N. Shang, “Privacy-preserving Digital Identity Management for Cloud Computing,” IEEE Data Eng. Bull, vol. 32, no. 1, pp. 21–27, 2009.

[4] I. Brandic, S. Dustdar, T. Anstett, D. Schumm, F. Leymann, and R. Konrad, “Compliant Cloud Computing (C3):

Architecture and Language Support for User-Driven Compliance Management in Clouds," in Proc. of CLOUD'10, 2010.

[5] Jingwei Huang and David M Nicol, Trust mechanisms for cloud computing, April 2013

[6] J. Huang and D. M. Nicol, "Trust Mechanisms for Cloud Computing," Journal of Cloud Computing, vol. 2, no. 1, pp. 1–14, 2013.

[7] Kai Hwang Deyi Li, Trusted Cloud Computing with Secure Resources and Data Coloring, Sept.-Oct. 2010

[8] K. Hoffman, D. Zage, and C. Nita-Rotaru, "A Survey of Attack and Defense Techniques for Reputation Systems," ACM Computing Surveys, vol. 42, no. 1, pp. 1–31, 2009.

[9] Lina Yao Quan Z. Sheng Zakaria Maamar, Achieving High Availability of Web Services Based on A Particle Filtering Approach, 2012

[10] R. Ko, P. Jagadpramana, M. Mowbray, S. Pearson, M. Kirchberg, Q. Liang, and B. Lee, "TrustCloud: A Framework for Accountability and Trust in Cloud Computing," in Proc. SERVICES'11, 2011.

[11] S. Habib, S. Ries, and M. Muhlhauser, "Towards a Trust Management System for Cloud Computing," in Proc. of TrustCom'11, 2011.

[12] Sheikh Mahbub Habib , Sebastian Ries y, Max M• uhlh• auser, Towards a Trust Management System for Cloud Computing

[13] Siani Pearson and Azzedine Benameur, Privacy, Security and Trust Issues Arising from Cloud Computing , 2010

[14] S. M. Khan and K. W. Hamlen, "Hatman: Intra-Cloud Trust Management for Hadoop," in Proc. CLOUD'12, 2012.

[15] S. Pearson, "Privacy, Security and Trust in Cloud Computing," in Privacy and Security for Cloud Computing, ser. Computer Communications and Networks, 2013, pp. 3–42.

[16] S. Pearson and A. Benameur, "Privacy, Security and Trust Issues Arising From Cloud Computing," in Proc. CloudCom'10, 2010.

[17] T. H. Noor, Q. Z. Sheng, and A. Alfazi, "Reputation Attacks Detection for Effective Trust Assessment of Cloud Services," in Proc. of TrustCom'13, 2013.

[18] T. H. Noor, Q. Z. Sheng, A. H. Ngu, A. Alfazi, and J. Law, "CloudArmor: A Platform for Credibility-based Trust Management of Cloud Services," in Proc. of CIKM'13, 2013.

[19] T. Noor and Q. Z. Sheng, "Credibility-Based Trust Management for Services in Cloud Environments," in Proc. of ICSOC'11, 2011.

[20] W. Conner, A. Iyengar, T. Mikalsen, I. Rouvellou, and K. Nahrst-edt, "A Trust Management Framework for Service-Oriented Environments," in Proc. of WWW'09, 2009.

ICADET: conference proceedings: 2024 Advanced Development in Engineering And Technology
ISSN: 2454-9924

Cost and Energy optimization for Big Data Processing in Geo-Spread Data Centers

M. Nithya #1, E. Indra *1

Mailam Engineering College, Mailam #1, *1

Nithyasundari5@gmail.com #1

Abstract

The high volume of demands on big data processing produces a heavy load on calculation, storage, and communication in data storage, which hence determines the required operational measures to data storage. Data center resizing (DCR) has been proposed to reduce the computation cost by adjusting the number of activated servers via task placement. MINLP (Mixed Integer Non Linear Programming) is the problem of non-joint optimization. MILP (Mixed Integer Linear Programming) is the problem of joint optimization of Task Assignment, Data Placement, and Data Movement. Markov chain is used to derive the execution time of data centers.

Keywords – Demand, Big data, Measures, Linear Programming

1 Introduction

Data explosion leads to demand for big data processing in data centers that are distributed at different geographic regions. Data computation, storage, and communication in data centers, which hence incurs considerable operational expenditure to data center providers. Computation tasks conducted only when the corresponding data is available due to tight coupling between data .Task assignment, data placement and data movement, deeply influence the operational expenditure of data centers. Many efforts have been made to lower the computation or communication cost of data centers. Data center resizing (DCR) has been proposed to reduce the computation cost by adjusting the number of activated servers via task placement. Based on DCR, some studies have explored the geographical distribution nature of data centers and electricity price heterogeneity to lower the electricity cost. Big data service frameworks, e.g., comprise a distributed file system underneath, which distributes data chunks and their replicas across the data centers for fine-grained load-balancing and high parallel data access performance. To reduce the communication cost, a few recent studies make efforts to improve data locality by placing jobs on the servers where the input data reside to avoid remote data

loading. Although the above solutions have obtained some positive results, they are far from achieving the cost efficient big data processing because of the following weaknesses. First, data locality may result in a waste of resources. For example, most computation resource of a server with less popular data may stay idle. The low resource utility further causes more servers to be activated and hence higher operating cost. Second, the links in networks vary on the transmission rates and costs according to their unique features, e.g. the distances and physical optical fiber facilities between data centers. However, the existing routing strategy among data centers fails to exploit the link diversity of data center networks. Due to the storage and computation capacity constraints, not all tasks can be placed onto the same server, on which their corresponding data reside. It is unavoidable that certain data must be downloaded from a remote server. In this case, routing strategy matters on the transmission cost, the transmission cost, e.g., energy, nearly proportional to the number of network link used. The more link used, the higher cost will be incurred. Therefore, it is essential to lower the number of links used while satisfying all the transmission requirements. Third, the Quality-of-Service (QoS) of big data tasks has not been considered in existing work. Similar to conventional cloud services, big data applications also exhibit Service-Level-Agreement (SLA) between a service provider and the requesters. To observe SLA, a certain level of QoS, usually in terms of task completion time, shall be guaranteed. The QoS of any cloud computing tasks is first determined by where they are placed and how many computation resources are allocated.

Besides, the transmission rate is another influential factor since big data tasks are data-centric and the computation task cannot proceed until the corresponding data are available. Existing studies, e.g., on general cloud computing tasks mainly focus on the computation capacity constraints, while ignoring the constraints of transmission rate.

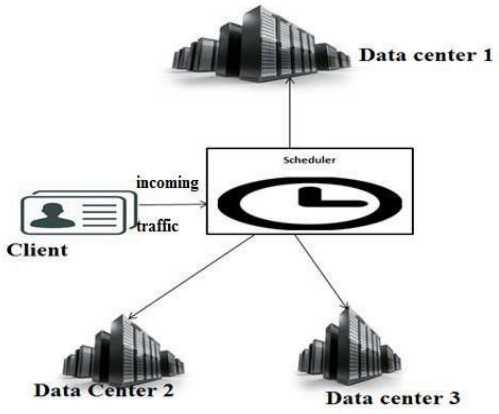
2 Related Works

In 2015, 71% of worldwide data center hardware spending will come from the big data processing, it's predicted by Gartner. In Data center resizing (DCR) data locality may result in a waste of resources. Less popular data may stay idle and the low resource utility causes more servers to be activated and hence higher operating cost. Links in networks vary on the transmission rates and costs according to their unique features. If the routing strategy among data centers fails then it is unavoidable to download from a remote server. In this case, routing strategy matters on the transmission cost. The Quality-of-Service (QoS) of big data tasks has not been considered in existing work. Big data processing translated into big price due to its high demand on computation, communication resources. Decrease the system reliability by continuous processing. Consumes high energy for high computation process its cause negative impacts to

environment. In our proposed mechanism, formulate the cost minimization problem based on the closed-form expression in a form of mixed integer nonlinear programming (MINLP). Linearize it as a mixed-integer linear programming (MILP) problem for solving the complexity of MINLP. Cost for high computational data is minimized. Reduce the system operation increases system reliability. Energy Consumption is minimized.

3 System Design

In this section, we introduce the system model. From this architecture, determines with the following components, Data center1, Data Center 2 and Data center 3. Also determines with Client and the scheduler process



4 Methodologies

4.1 Uploading of data in big data

Select the big data and stored into the hadoop environment for performing map reduce on hadoop. The data should be loaded into the VM server location. After uploading the file the data segmentation is performed for further process.

4.2 Packet Separation

Packet segmentation improves network performance by splitting the packets in received Ethernet frames into separate buffers. Packet segmentation may be responsible for splitting one into multiple so that reliable transmission of each one can be performed individually. The packet processing system is specifically designed for dealing with the network traffic. Segmentation may be required when the

data packet is larger than the maximum transmission unit supported by the network.

4.3 Job Responsibility

The Data Center should be selected according to computation and storage capacity of servers resides in the data center. Identification of Data Center is important matter for minimizing operational expenditure of servers reside in the each data centers. Data chunks can be placed in the same data center when more servers are provided in each data center. Further increasing the number of servers will not affect the distributions of tasks. Task is assigned to data center according to Memory requirement for effectively processing of data.

4.4 Loading of data

A Data Placement on the servers and the amount of load capacity assigned to each file copy so as to minimize the communication cost while ensuring the user experience. Optimization scheme that simultaneously optimize the virtual machine (VM) placement and network flow routing to maximize energy savings.

4.5 Assessment Process

We present the performance results of our joint-optimization algorithm using the MILP formulation. Evaluate server cost, communication cost and overall cost under different total server numbers.

5 Algorithm Description

Mixed Integer Non Linear Programming is the Non Joint Optimization problem. Mixed-integer optimization provides a powerful framework for mathematically modeling many optimization problems that involve discrete and continuous variables. The important factor for minimizing cost is Task Assignment, Data Loading and Data Movement. Optimization of these three factors will reduce the overall cost of network. The Non Joint Optimization technique individually optimizes these factors. So time increasing will lead to increase of operational cost of overall network. To reduce the System complexity (Continuous processing) of overall network is causing the unreliability of system. To reduce the system complexity of overall network, we linearize the MINLP by changing some operational parameters of MILP. MILP is a Joint Optimization of Task Assignment, Data Loading and Data Movement. Branch and Bound, Outer-Approximation, Generalized Benders and Extended Cutting Plane methods, as applied to nonlinear discrete optimization problems that are expressed in algebraic form. The solution of MINLP problems with convex functions is presented first, followed by a brief discussion on extensions for the non-convex case. Properties of the algorithms are first considered for the case when the nonlinear functions are convex in the discrete and continuous variables. Extensions are then presented for handling nonlinear equations and non-convexities.

6 Literature Review

[21] Describe about, Computing equipment can be safely and efficiently hosted within a given power budget. Load variation and statistical effects are the main dynamic sources of inefficiency in power deployment. Large-scale Internet services require a computing infrastructure that can be appropriately described as a warehouse-sized computing system. The cost of building datacenter facilities capable of delivering a given power capacity to such a computer can rival the recurring energy consumption costs themselves. Therefore, there are strong economic incentives to operate facilities as close as possible to maximum capacity, so that the non-recurring facility costs can be best amortized. That is difficult to achieve in practice because of uncertainties in equipment power ratings and because power consumption tends to vary significantly with the actual computing activity. Effective power provisioning strategies are needed to determine how much computing equipment can be safely and efficiently hosted within a given power budget. In this paper we present the aggregate power usage characteristics of large collections of servers (up to 15 thousand) for different classes of applications over a period of approximately six months. Those observations allow us to evaluate opportunities for maximizing the use of the deployed power capacity of datacenters, and assess the risks of over-subscribing it. We find that even in well-tuned applications there is a noticeable gap (7 - 16%) between achieved and theoretical aggregate peak power usage at the cluster level (thousands of servers). The gap grows to almost 40% in whole datacenters. This headroom can be used to deploy additional compute equipment within the same power budget with minimal risk of exceeding it. We use our modeling framework to estimate the potential of power management schemes to reduce peak power and energy usage. We find that the opportunities for power and energy savings are significant, but greater at the cluster-level (thousands of servers) than at the rack-level (tens). Finally we argue that systems need to be power efficient across the activity range, and not only at peak performance levels. [9] Describe about, Video-on-Demand (VoD) services require frequent updates in file configuration on the storage subsystem, so as to keep up with the frequent changes in movie popularity. This defines a natural reconfiguration problem in which the goal is to minimize the cost of moving from one file configuration to another. The cost is incurred by file replications performed throughout the transition. The problem shows up also in production planning, preemptive scheduling with set-up costs, and dynamic placement of Web applications. We show that the reconfiguration problem is NP-hard already on very restricted instances. We then develop algorithms which achieve the optimal cost by using servers whose load capacities are increased by O(1), in particular, by factor $1 + \delta$ for any small $0 < \delta < 1$ when the number of servers is fixed, and by factor of $2 + \varepsilon$ for arbitrary number of servers, for some $\varepsilon \in [0, 1)$. To the best of our knowledge, this particular variant of the

data migration problem is studied here for the first time. [3] Describe about, In light of the challenges of effectively managing Big Data, we are witnessing a gradual shift towards the increasingly popular Linked Open Data (LOD) paradigm. LOD aims to impose a machine-readable semantic layer over structured as well as unstructured data and hence automate some data analysis tasks that are not designed for computers. The convergence of Big Data and LOD is, however, not straightforward: the semantic layer of LOD and the Big Data large scale storage do not get along easily. Meanwhile, the sheer data size envisioned by Big Data denies certain computationally expensive semantic technologies, rendering the latter much less efficient than their performance on relatively small data sets. In this paper, we propose a mechanism allowing LOD to take advantage of existing large-scale data stores while sustaining its "semantic" nature. We demonstrate how RDF-based semantic models can be distributed across multiple storage servers and we examine how a fundamental semantic operation can be tuned to meet the requirements on distributed and parallel data processing. Our future work will focus on stress test of the platform in the magnitude of tens of billions of triples, as well as comparative studies in usability and performance against similar offerings.

7 Conclusion

From this Cost and Energy optimization for Big Data Processing in Geo-Spread Data Centers has been implemented. To study the data placement, task assignment, data center resizing and routing for minimize the overall operational cost. Characterize the data processing process using a two-dimensional Markov chain Derive the expected completion time in closed-form**.** To tackle the high computational complexity of solving our MINLP, we linearize it into an MILP problem. Additionally in future, Fault tolerance mechanisms either consume significant extra energy to detect and recover from the failures. Fault-tolerant describes a computer system or component designed so that, in the event that a component fails, a backup component or procedure can immediately take its place with no loss of service.

8 References

[1] A. Cidon, R. Stutsman, S. Rumble, S. Katti, J. Ousterhout, and M. Rosenblum, “MinCopysets: Derandomizing Replication In Cloud Storage,” in The 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI), 2013.

[2] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, “Cutting the Electric Bill for Internet-scale Systems,” in Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM). ACM, 2009, pp. 123–134.

[3] B. Hu, N. Carvalho, L. Laera, and T. Matsutsuka, towards big linked data: a large-scale, distributed semantic data storage, November 2014

[4] B. L. Hong Xu, Chen Feng, “Temperature Aware Workload Management in Geo-distributed

Datacenters," in Proceedings of International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS). ACM, 2013, pp. 33–36.

[5] "Data Center Locations," http://www.google.com/about/datacenters/inside/locations/index.html.

[6] F. Chen, M. Kodialam, and T. V. Lakshman, "Joint scheduling ofprocessing and shuffle phases in mapreduce systems," in Proceedings of the 29th International Conference on Computer Communications (INFOCOM). IEEE, 2012, pp. 1143–1151.

[7] "Gurobi," www.gurobi.com.

[8] H. Jin, T. Cheocherngngarn, D. Levy, A. Smith, D. Pan, J. Liu, and N. Pissinou, "Joint Host-Network Optimization for EnergyEfficient Data Center Networking," in Proceedings of the 27th International Symposium on Parallel Distributed Processing (IPDPS), 2013, pp. 623–634.

[9] H. Shachnai, G. Tamir, and T. Tamir, Minimal Cost Reconfiguration of Data Placement in a Storage Area Network, September 2009.

[10] I. Marshall and C. Roadknight, "Linking cache performance to user behaviour," Computer Networks and ISDN Systems, vol. 30, no. 223, pp. 2123 – 2130, 1998.

[11] J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton, "Mad skills: new analysis practices for big data," Proc. VLDB Endow., vol. 2, no. 2, pp. 1481–1492, 2009. L. Kleinrock, "The latency/bandwidth tradeoff in gigabit networks," Communications Magazine, IEEE, vol. 30, no. 4, pp. 36–40, 1992.

[12] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," Communications of the ACM, vol. 51, no. 1, pp. 107–113, 2008.

[13] L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing Electricity Cost: Optimization of Distributed Internet Data Centers in a MultiElectricity-Market Environment," in Proceedings of the 29th International Conference on Computer Communications (INFOCOM). IEEE, 2010, pp. 1–9.

[14] M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, and D. Borthakur, "Xoring elephants: novel erasure codes for big data," in Proceedings of the 39th international conference on Very Large Data Bases, ser. PVLDB'13. VLDB Endowment, 2013, pp. 325–336

[15] P. X. Gao, A. R. Curtis, B. Wong, and S. Keshav, "It's Not Easy

Being Green," in Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM). ACM, 2012, pp. 211–222.

[16] R. Kaushik and K. Nahrstedt, "T*: A data-centric cooling energy

costs reduction approach for Big Data analytics cloud," in 2012 International Conference for High Performance Computing, Networking, Storage and Analysis (SC), 2012, pp. 1–11.

[17] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu, "No "Power" Struggles: Coordinated Multi-level Power Management for the Data Center," in Proceedings of the 13th x ACM, 2008, pp. 48–59.

[18] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam, "Optimal Power Cost Management Using Stored Energy in Data Centers," in Proceedings of International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS). ACM, 2011, pp. 221–232.

[19] S. A. Yazd, S. Venkatesan, and N. Mittal, "Boosting energy efficiency with mirrored data block replication policy and energy scheduler," SIGOPS Oper. Syst. Rev., vol. 47, no. 2, pp. 33–40, 2013.

[20] S. Govindan, A. Sivasubramaniam, and B. Urgaonkar, "Benefits and Limitations of Tapping Into Stored Energy for Datacenters,"in Proceedings of the 38th Annual International Symposium on Computer Architecture (ISCA). ACM, 2011, pp. 341–352.

[21] Xiaobo Fan, Wolf-Dietrich, Weber Luiz André Barroso., Power Provisioning for a Warehouse-sized Computer, June 2007

ISSN: 2454-9924

DATA HIDING AND COMPRESSION METHOD FOR DIGITALIMAGES USING SIDE MATCH VECTOR QUANTIZATION

S.Yogalakshmi
M.Tech(ECE)
Christ College of Engineering and Technology.

T.Somassoundaram
Research Scholar
Sathyabama University

ABSTRACT:

In the emerging technology people used to share and transmit digital content with each other conveniently. In order to guarantee communication efficiently and save the network bandwidth, data hiding and compression technique are implemented jointly in a single module. Today reversible data hiding is considered as one of the latest research area in the field of secret data hiding technique. Basically data hiding in our proposal is applied in such a way that the secret data are hidden in some cover images such as audio, video, image, etc. This paper covers mainly on data hiding and compression techniques like vector quantization and side match vector quantization.

***Index terms*:**

Data hiding, image compression, vector quantization, side match vector quantization.

I. INTRODUCTION:

Data hiding plays an important role in the field of information security. This technique can be used to prevent the transmitted content from the impending attraction of malicious attackers. As a result, the privacy of secret information is maintained. The use of data hiding techniques is altered from that of traditional cryptography or watermarking techniques. The role of cryptography is to encrypt the message into a meaningless data in such a way that it should not be attacked by any foreign agent. Beside cryptographic method the use of watermarking is to protect the copy right.

There are mainly two types of data hiding techniques. They are

- Reversible data hiding.
- Irreversible data hiding.

In reversible data hiding method both the secret message and the cover media are recovered completely, but in irreversible data hiding method only secret message are recovered. Some of the two important uses of data hiding in digital media are to provide the proof of the copyright and assurance of content integrity. Another application includes the inclusion of argumentation of data.

With a rapid development of internet technology, people used to convey their data and sent them in such a way that they are effectively utilized in order the scrambling problem are reduced. Image compression techniques reduce the redundancy and irrelevance of the image pixels in order to able to store or transmit information in an efficient manner.

Two different types of compression techniques are as follows. They are

- Lossy compression technique
- Lossless compression technique

Lossy compression techniques creates smaller image by discarding excess image pixel from the original image. Whereas in lossless compression technique, it never removes any

pixels from the original image instead data's are represented in a mathematical formula.

which are applied to various compression techniques of digital images, such as JPEG, JPEG 2000 and vector quantization. Generally vector quantization is considered one of the most simplest and popular lossy compression method, because due its plainness and cost effectiveness in implementation. During the compression process of vector quantization, Euclidean distance is utilized to evaluate the similarity between each image block and codebooks are used for assigning code word in each image block. While moving to the decompression process only a simple lookup table operation is required for retrieving the corresponding index values in each block. The rest of the paper is organized as follows. Section II describes the existing studies. Proposed work is given in section III and section IV concludes the experimental results and conclusion in section V.

II. EXISTING STUDIES

Among many image compression technique vector quantization is one of the most accepted method. During the year of 2003, Du and Hsu [3] projected an adaptive data hiding method for vector quantization compressed images, in this method the process of embedding is varied based upon the amount of data hidden in it. In this method the codebook was sub divided into two or more sub codebook, and the best match is used for hiding the secret data. Later this method is found to have a low embedding capacity. In order to improve its embedding rate, a VQ-based data-hiding scheme by a code word clustering [4] technique was proposed. Here the secret data were embedded into the VQ index table by code word-order-cycle permutation. In this technique, more possibilities and flexibility can be offered to improve its performance. In 2009 Lin and Chen [5] adjusted the pre-determined distance threshold according to the required hiding capacity and arranged a number of similar code words in one group to embed the secret sub-message. The search-order coding

ISSN: 2454-9924

Presently, different types of data hiding schemes for the compressed codes has been reported

(SOC) algorithm was proposed by Hsieh and Tsai [6] , which can be utilized to further compress the VQ index table and achieve better performance of the bit rate through searching nearby identical image blocks following a spiral path. Some steganographic schemes were also proposed to embed secret data.

However in all the above schemes, data hiding and compression are performed separately in a single module. Under this condition the foreign users may have an opportunity to intercept the compressed image and which leads to lower efficiency in many applications.

III. PROPOSED WORK

In order to overcome the drawbacks of existing studies, the proposed method establishes a joint data hiding and compression process. Here the process of data hiding and compression are found to be integrated in a single module which avoids the risks of foreign attackers and increases its implementation efficiency. The proposed method is based on the use of side match vector quantization.

A. Image Compression Technique

In our scheme, the sender and the receiver both have the same codebook with W code words, and each code word length is n^2. Denote the original uncompressed image sized $M \times N$ as **I**, and it is divided into the non-overlapping $n \times n$ blocks. For simplicity, we assume that M and N can be divided by n with no remainder. Denote all k divided blocks in raster scanning order as **B***i*, *j*, where $k = M \times N/ n2$, $i = 1,2, . , M/n$ and j
= 1,2, .*N/n*. Before being embedded, the secret bits are scrambled by a secret key to ensure security. The blocks in the leftmost and topmost of the image **I**, i.e.,**B***i*,1*(i* = 1, 2, . . . , *M/n)* and **B**1, *j (j* = 2, 3, . . . , *N/n),* are encoded by VQ directly and are not used to embed secret bits. The residual blocks are encoded progressively in raster scanning order, and their encoded

methods are related to the secret bits for embedding.

B. Secret Data Embedding Algorithm

The process of secret data embedding into the JPEG compressed image includes the following steps.

1. Get an image and apply the process of entropy decoding.
2. Let F be the quantized DCT block with $F(i, j)$. Embed the secret data with length $E(i, j)$ in the LSBs of $F(i, j)$.
3. The process of data hiding and compression follows the raster scanning order to embed the secret bit.
4. The blocks are chosen based on L-shape pattern.
5. After the blocks are chosen the secret data's are embedded based on the threshold value. The process gets stopped after all blocks are embedded by secret data.

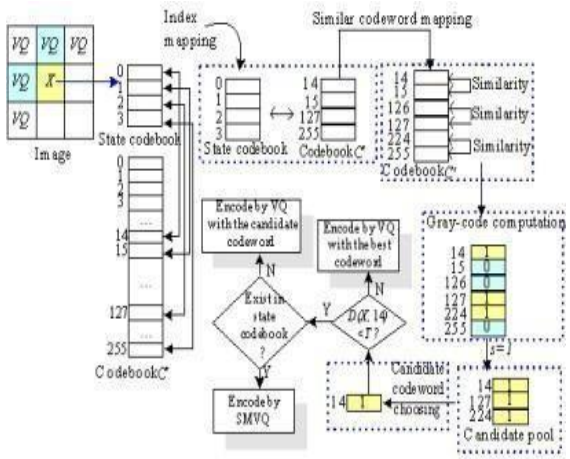


Fig.1 Flow chart of the data embedding phase

C. Data Extracting Algorithm

The process of secret data extraction includes the following steps

1. Let the currently processed (decoded) image block be *xi*, extract the first bit from the received bit stream.
2. If the extracted indicator bit is equal to 0, read the index value with log w bits and decompress it by using VQ and extract the watermark bit.
3. Otherwise if the extracted indicator bit is equal to 1, read the index value with

log (R+1) bits and decompress it by using SMVQ and extract the watermark bit.

4. The process is repeated to the overall bit stream and the corresponding secret data are extracted.

The figure 2 shows the data extracting phase.

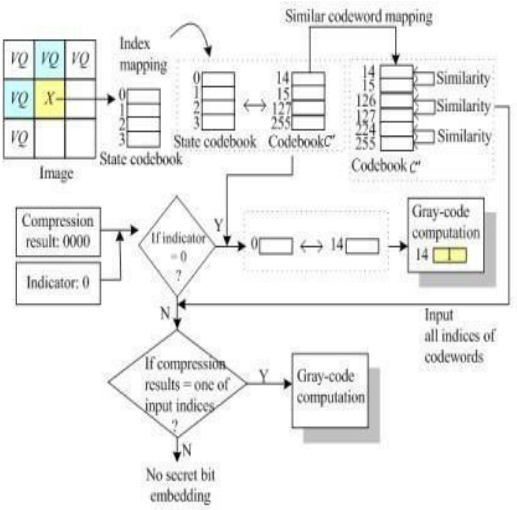


Fig.2 Flow chart of the data extraction phase

IV. EXPERIMENTAL RESULTS

Experiments were conducted on a group of gray-level images to verify the effectiveness of the scheme. In this experiment, the sizes of the images were divided into non-overlapping image blocks i.e., $n = 4$. Accordingly, the length of each codeword in the VQ codebooks was 16. The parameter P was set to 15. Six standard, 512 × 512 test images are shown in figure 3 i.e. Lena, Airplane, Lake, Peppers, Sailboat, and Tiffany, are shown in. Apart from this six standard images, the uncompressed color image database (UCID) also contains 1338 various color images with sizes of 512 × 384 was also adopted. The luminance components of the color images in this database were used in the experiments. The performance of compression ratio, decompression quality, and hiding capacity for the proposed scheme were evaluated. All experiments were implemented on a computer with Windows 7 operating system, and the programming software was Mat lab.

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN : 2454-9924

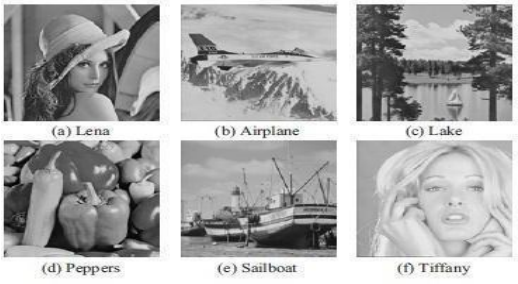


Fig. 3 Six standard test images.

In this scheme, the hiding capacity and the visual quality of cover images are mainly affected by the three parameters, the variance threshold *THvr*, the side match distortion threshold *THsmvd*, and the side-match state code book size *p*. These parameters are familiar based on the amount of data hidden and the characteristics of the cover image. They can be used as keys for the extraction of secret data. If *THvr* is set with a well-built value, more blocks will be treated as even blocks and, consequently, more secret data can be unseen into a cover image. However, the visual quality of cover image will be degraded, since more blocks were directly predicted by the proposed scheme. If *THsmvd* is given as a well-built value, more even blocks will be selected for hiding data. Therefore, the hiding capacity increases and the visual quality is reduced for the cover image. If *p* is assigned to be a well-built value, more code words are included into the state codebook and the selected even blocks will be encoded (predicted) more randomly. Accordingly, the visual quality of cover image degrades while the hiding capacity increases.

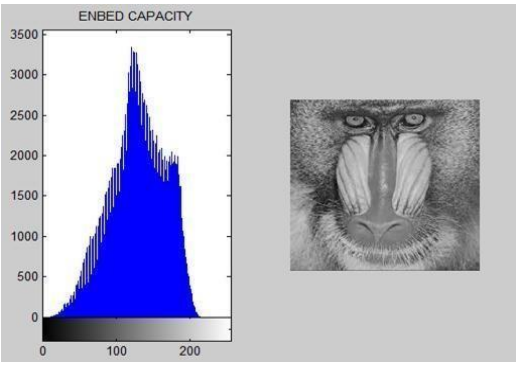


Fig.4 Image representing hidden data and histogram

Besides, we employed the peak signal-to-noise ratio (PSNR) as a measure of the stegno-image quality. It is defined as follows:

$$\mathrm{PSNR} = 10 \times \log_{10} \frac{255^2}{MSE} \mathrm{dB}$$

Where,*MSE* is the mean-square error. For an $N \times N$ image, its *MSE* is defined as,

$$MSE = \left(\frac{1}{N}\right)^2 \times \sum_{i=1}^{N} \sum_{j=1}^{N} (x[i,j] - \bar{x}[i,j])^2.$$

Here, $x[i, j]$ and $\bar{x}[i, j]$ denote the original and decoded gray levels of the pixel $[i, j]$ in the image, respectively. A well-built PSNR value means that the stegno-image preserves the original image quality better. Our method employs the capacity factors α to control the level of embedding capacity. Users can adjust it to balance between the image quality (PSNR) and the embedding capacity. If the capacity factor is selected as a large number, then the embedding capacity can be raised, but the cost is that the compression ratio of the image gets low. Through quite a number of experiments, the capacity factor α is finally selected for uniform blocks, and $1.1 \times \alpha$ for non-uniform blocks.

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN :

2454-9924

V. CONCLUSION

The main contribution of the proposed method is to improve the data hiding capacity. Our method embeds a joint data-hiding and compression scheme by using SMVQ. The blocks, except in the leftmost and topmost of the image, can be embedded with secret data and compressed simultaneously, and the adopted compression method switches between SMVQ adaptively according to the embedding bits. VQ is also utilized for some complex blocks to control the visual distortion and error diffusion. On the receiver side, after segmenting the compressed codes into a series of sections by the indicator bits, the embedded secret bits can be easily extracted according to the index values in the segmented sections, and the decompression for all blocks can also be achieved successfully by VQ, SMVQ. Ours is an adaptive data hiding method with which one can adjust capacity factor to balance between the image quality and the embedding capacity dynamically.

REFERENCES

1. Chuan Qin, Chin-Chen Chang, *Fellow, IEEE, and Yi-Ping Chiu* , "A Novel Joint Data Hiding and Compression Scheme Based on SMVQ and Image Inpainting," *IEEE Trans. On Image Processing,* Vol. 23, No. 3, March 2014.
2. W. J. Wang, C. T. Huang, and S. J. Wang, "*VQ applications in steganographic data hiding upon multimedia images*," *IEEE Syst. J.*, vol. 5, no. 4, pp. 528–537, Dec. 2011.
3. W. C. Du and W. J. Hsu, "*Adaptive data hiding based on VQ compressed images*," *IEE Proc. Vis., Image Signal Process.*, vol. 150, no. 4, pp. 233–238, Aug. 2003.
4. C. C. Chang and W. C. Wu, "*Hiding secret data adaptively in vector quantisation index tables*," *IEE Proc. Vis., Image Signal Process.*, vol. 153, no. 5, pp. 589–597, Oct. 2006.
5. C. C. Lin, S. C. Chen, and N. L. Hsueh, "Adaptive *embedding techniques for VQ-compressed images," Inf. Sci.*, vol. 179, no. 3, pp. 140–149, 2009.
6. C. H. Hsieh and J. C. Tsai, "*Lossless compression of VQ index with search-order coding*," *IEEE Trans. Image Process.*, vol. 5, no. 11, pp. 1579–1582, Nov. 1996.
7. H. W. Tseng and C. C. Chang, "*High capacity data hiding in JPEG compressed images*, "*Informatica,* vol. 15, no. 1, pp. 127–142, 2004.
8. Y. C. Hu, "*High-capacity image hiding scheme based on vector quantization*," *Pattern Recognit.*, vol. 39, no. 9, pp. 1715–1724, 2006.
9. Y. P. Hsieh, C. C. Chang, and L. J. Liu, "*A two-codebook combination and three-phase block matching based image-hiding scheme with high embedding capacity*," *Pattern Recognit.*, vol. 41, no. 10, pp. 3104–3113,2008.
10. H. Yang and Y. C. Lin, "*Fractal curves to improve the reversible data embedding for VQ-indexes based on locally adaptive coding*," *J. Vis.Commun. Image Represent.*, vol. 21, no. 4, pp. 334–342, 2010.

ISSN: 2454-9924

Detection of Look Alike Detection of Clone Node and Collusion Attacks in WSN

R. Aswini #1, P. R. Jayanthi *1

Mailam Enginnering College, Mailam #1, *1

Raswini29@gmail.com #1

Abstract - The combination of node is naturally accomplished due to computational power and energy resources. In the previous mechanism, Wireless sensor networks are vulnerable to the node clone, and several distributed protocols have been proposed to detect this attack. So they require too strong assumptions to be practical for large-scale, randomly deployed sensor networks. In the proposed mechanism, we use two new node clone detection protocols with different agreement on network conditions and performance. The first one is based on a Distributed Hash Table (DHT) in which Chord algorithm is used to detect the cloned node, every node is assigned with the primary key, and before it transfer the data it has to give its key which would be verified by the proof node. If same key is given by another Node then the proof node detects the cloned Node. The second one is based on the Distributed Detection Protocol which is same as DHT, but it is easy and cheaper determination. Here every node only needs to know the neighbor-list containing all neighbor IDs and its locations. In the modified Process, we are determining RDE protocol, by location based nodes detection, where every region/location will have a group leader. The Group leader will generate a random number with time stamp to the available nodes in that location. Witness nodes verify the random number and time stamp to detect the cloned node. Here, it also provides security measures.

Key Words – Computational Power, Energy resources, Distributed Hash Table, Time Stamp

1 Introduction

DUE to a need for robustness of monitoring and low cost of the nodes, wireless sensor networks (WSNs) are usually redundant. Data from multiple sensors is aggregated at an aggregator node which then forwards to the base station only the aggregate values. At present, due to limitations of the

computing power and energy resource of sensor nodes, data is aggregated by extremely simple algorithms such as averaging. However, such aggregation is known to be very vulnerable to faults, and more importantly, malicious attacks. This cannot be remedied by cryptographic methods, because the attackers generally gain complete access to information stored in the compromised nodes. For that reason data aggregation at the aggregator node has to be accompanied by an assessment of trustworthiness of data from individual sensor nodes. Thus, better, more sophisticated algorithms are needed for data aggregation in the future WSN. Such an algorithm should have two features. 1. In the presence of stochastic errors such algorithm should produce estimates which are close to the optimal ones in information theoretic sense. Thus, for example, if the noise present in each sensor is a Gaussian independently distributed noise with zero mean, then the estimate produced by such an algorithm should have a variance close to the Cramer- Rao lower bound (CRLB), i.e, it should be close to the variance of the Maximum Likelihood Estimator (MLE). However, such estimation should be achieved without supplying to the algorithm the variances of the sensors, unavailable in practice. 2. The algorithm should also be robust in the presence of non-stochastic errors, such as faults and malicious attacks, and, besides aggregating data; such algorithm should also provide an assessment of the reliability and trustworthiness of the data received from each sensor node. Trust and reputation systems have a significant role in supporting operation of a wide range of distributed systems, from wireless sensor networks and e-commerce infrastructure to social networks, by providing an assessment of trustworthiness of participants in such distributed systems. A trustworthiness assessment at any given moment represents an aggregate of the behavior of the participants up to that moment and has to be robust in the presence of various types of faults and malicious behavior. There are a number of incentives for attackers to manipulate the trust and reputation scores of participants in a distributed system, and such manipulation can severely impair the performance of such a system. The main target of malicious attackers is aggregation algorithms of trust and reputation systems.

2 Related work

In the previous mechanism, Wireless sensor networks are vulnerable to the node clone,

and several distributed protocols have been proposed to detect this attack. So they require too strong assumptions to be practical for large-scale, randomly deployed sensor networks. Here it provides some, drawbacks are, Less Security, Data hacking, missing privacy

2.1 Proposed Mechanism

In the proposed mechanism, we use two novel node clone detection protocols with different tradeoffs on network conditions and performance. The first one is based on a distributed hash table (DHT) in which Chord algorithm is used to detect the cloned node, every node is assigned with the unique key, and before it transmits the data it has to give its key which would be verified by the witness node. If same key is given by another Node then the witness node identifies the cloned Node. The second one is based on the Distributed Detection Protocol which is same as DHT, but it is easy and cheaper implementation. Here every node only needs to know the neighbor-list containing all neighbor IDs and its locations. In the modified system, Process, we are implementing RDE protocol, by location based nodes identification, where every region/location will have a group leader. The Group leader will generate a random number with time stamp to the available nodes in that location. Witness nodes verify the random number and time stamp to detect the cloned node. The message is also encrypted for security purpose. Here, it provides some advantages are, High security, Data integrity, easily find the attacker

3 Methodologies

3.1 Establishment of Network

This module is developed in order to create a dynamic network. In a network, nodes are interconnected with the admin, which is monitoring all the other nodes. All nodes are sharing their information with each other‟s

3.2 Distribution of Proof node

A major issue in designing a protocol to detect clone attacks is the selection of the witnesses. We will call „Witness‟ as a node that detects the existence of a node in two different locations within the same protocol run. If the adversary knows the future witnesses before the detection protocol executes, the adversary could subvert these nodes so that the attack goes undetected.

Here, we have identified two kinds of predictions:

1. ID-based prediction

2. Location-based prediction.

We say that a protocol for replica detection is ID-oblivious if the protocol does not provide any information on the ID of the sensors that will be the witnesses of the clone attack during the next protocol run. Similarly, a protocol is area-oblivious if probability does not depend on the geographical position of node in the network. Clearly, when a protocol is neither ID-oblivious nor area-oblivious, then a smart adversary can have good chances of succeeding, since it is able to use this information to subvert the nodes that, most probably, will be the witnesses.

3.3 Confirmation of Random Number

Random Key pre-distribution security scheme is implemented in the sensor network. That is, each node is assigned a number randomly with Time Stamp from Group Leader. Then the Group Leader will transmit Random Number (Encrypted with RSA algorithm) which was generated with respect to that Time Stamp to the Witness node. Witness node will now check the Random number which is generated with the User information. If both the data are matched then the Witness node will confirm that this node is Genuine.

3.4 Verification of User information

Each node is assigned an ID as individual once it is registered into the network and also an ID for the whole group (i.e.) Location ID is generated for each and every Location. That Node ID and Location ID are also appended with 1 (Encrypted with RSA algorithm). Then the Witness node will now check the node ID + Location ID which is generated with the User Information. If both the data are matched then the Witness node will confirm that this node with that Location is Genuine.

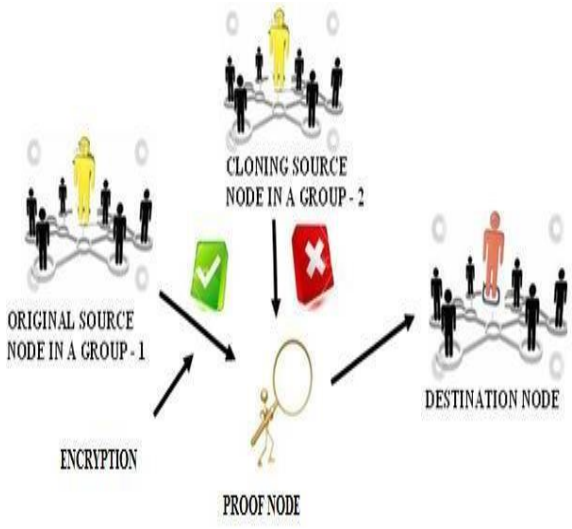
3.5 Replica Detection and Transfer

Only the Witness node confirms the Sender node, the data is send to the Destination, which is Genuine. If user specified information and the internal information are varied then the Witness node will identify that Cloning or some Mal

ISSN: 2454-9924

practice has occurred and the Packets are discarded by the witness node.

4 Architecture



From this architecture, it implemented with the following components, Original source node, cloning node, destination node, and the proof node. We provide a thorough empirical evaluation of effectiveness and efficiency of our proposed aggregation method. The results show that our method provides both higher accuracy and better collusion resistance than the existing methods.

5 Description of Algorithm

Input: a, b, c

Output: Estimation vector r

e→0, Q(0)→1;

Repeat

Calculate p(e+1)

Calculate f;

e→e+1;

Until estimation has processed

Here we assume that sensors are deployed in a hostile unattended environment. Consequently, some nodes can be physically compromised. We assume that when a sensor node is com- promised, all the information which is inside the node becomes accessible by the adversary. Thus, we cannot rely on cryptographic methods for preventing the attacks, since the adversary may extract cryptographic keys from the compromised nodes. We assume that through the compromised sensor nodes the adversary can send false data to the aggregator with a purpose of distorting the aggregate values. We also assume that all compromised nodes can be under control of

a single adversary or a colluding group of adversaries, enabling them to launch a sophisticated attack. We also consider that the adversary has enough knowledge about the aggregation algorithm and its parameters. Finally, we assume that the base station and aggregator nodes cannot be compromised in this adversary model; there is an extensive literature proposing how to deal with the problem of compromised aggregators; in this paper we limit our attention to the lower layer problem of false data being sent to the aggregator by compromised individual sensor nodes, which has received much less attention in the existing literature.

6 Conclusion

From this Detection of Look Alike Detection of Clone Node and Collusion Attacks in WSN have been implemented, here we introduced a novel collusion attack scenario against a number of existing IF algorithms. Moreover, we proposed an improvement for the IF algorithms by providing an initial approximation of the trustworthiness of sensor nodes which makes the algorithms not only collusion robust, but also more accurate and faster converging. In future work, we will investigate whether our approach can protect against compromised aggregators. We also plan to implement our approach in a deployed sensor network.

7 References

[1] A. Jøsang and J. Golbeck, “Challenges for robust trust and reputation systems,” in Proc. 5th Int. Workshop Security Trust Manage., Saint Malo, France, 2009, pp. 253–262.

[2] B.-C. Chen, J. Guo, B. Tseng, and J. Yang, “User reputation in a comment rating environment,” in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 211, pp. 159–167.

[3] C. de Kerchove and P. Van Dooren, “Iterative filtering in reputation systems,” SIAM J. Matrix Anal. Appl., vol. 31, no. 4, pp. 1812– 1834, Mar. 2010.

[4] C. T. Chou, A. Ignatovic, and W. Hu, “Efficient computation of robust average of compressive sensing data in wireless sensor networks in the presence of sensor faults,” IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 8, pp. 1525–1534, Aug. 2013.

[5] D. Wagner, “Resilient aggregation in sensor networks,” in Proc. 2nd ACM

Workshop Security Ad Hoc Sens. Netw., 2004, pp. 78–87.

[6] E. Ayday, H. Lee, and F. Fekri, "An iterative algorithm for trust and reputation management," Proc. IEEE Int. Conf. Symp. Inf. Theory, vol. 3, 2009, pp. 2051–2055.

[7] H. Liao, G. Cimini, and M. Medo, "Measuring quality, reputation and trust in online communities," in Proc. 20th Int. Conf. Found. Intell. Syst., Aug. 2012, pp. 405–414.

[8] H.-L. Shi, K. M. Hou, H. ying Zhou, and X. Liu, "Energy efficient
and fault tolerant multicore wireless sensor network: E MWSN," in Proc. 7th Int. Conf. Wireless Commun., Netw. Mobile Comput.,

[9] H.-S. Lim, G. Ghinita, E. Bertino, and M. Kantarcioglu, "A gametheoretic approach for high-assurance of data trustworthiness in sensor networks," in Proc. IEEE 28th Int. Conf. Data Eng., Apr. 2012, pp. 1192–1203.

[10] H.-S. Lim, Y.-S. Moon, and E. Bertino, "Provenance-based trustworthiness assessment in sensor networks," in Proc. 7th Int. Workshop Data Manage. Sensor Netw., 2010, pp. 2–7. 2011, pp. 1–4.

11] L. Wasserman, All of Statistics : A Concise Course in Statistical Inference. New York, NY, USA: Springer,.

[12] K. Hoffman, D. Zage, and C. Nita-Rotaru, "A survey of attack and defense techniques for reputation systems," ACM Comput. Surveys, vol. 42, no. 1, pp. 1:1–1:31, Dec. 2009.

[13] M. Rezvani, A. Ignjatovic, E. Bertino, and S. Jha, "Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks," School Comput. Sci. and Eng., Univ. New South Wales, Kensington, NSW, Australia, Tech. Rep. UNSW-CSE-TR-201319, Jul. 2013.

[14] P. Laureti, L. Moret, Y.-C. Zhang, and Y.-K. Yu, "Information filtering via iterative refinement," Europhys. Lett., vol. 75, pp. 1006– 1012, Sep. 2006.

[15] R.-H. Li, J. X. Yu, X. Huang, and H. Cheng, "Robust reputation based ranking on bipartite rating networks," in Proc. SIAM Int. Conf. Data Mining, 2012, pp. 612–623.

[16] R. Roman, C. Fernandez-Gago, J. Lopez, and H. H. Chen, “Trust and reputation systems for wireless sensor networks,” in Security and Privacy in Mobile and Wireless Networking,S. Gritzalis, T. Karygiannis, and C. Skianis, eds.,Leicester, U.K.: Troubador Publishing Ltd, 2009 pp. 105–128,.

[17] S. Ozdemir and Y. Xiao, “Secure data aggregation in wireless sen- sor networks: A comprehensive overview,” Comput. Netw., vol. 53, no. 12, pp. 2022–2037, Aug. 2009.

[18] Y.-K. Yu, Y.-C. Zhang, P. Laureti, and L. Moret, “Decoding information from noisy, redundant, and intentionally distorted sources,” Physica A: Statist. Mech. Appl., vol. 371, pp. 732–744, Nov. 2006.

[19] Y. Yu, K. Li, W. Zhou, and P. Li, “Trust mechanisms in wireless sensor networks: Attack analysis and countermeasures,” J. Netw. Comput. Appl., vol. 35, no. 3, pp. 867–880, 2012. 2

[20] Y. Zhou, T. Lei, and T. Zhou, “A robust ranking algorithm to spamming,” Europhys. Lett., vol. 94, p. 48002, 2011.

ENHANCEMENT OF POWER QUALITY USING UNIFIED POWER QUALITY CONDITIONER

ANANTHAN.N
PG Scholar, Power system engineering, Arunai College of engineering, Anna University.
jananthan1991@gmail.com

ABSTRACT

In the modern power system the usage of power electronics loads are relatively high and it behaves as non-liner load. This load causes the serious voltage distortion and power quality issues on the transmission and distribution system by injecting the harmonics. The active power filters is used to regulate this problem. Unified power quality conditioner is the combination of series and shunt active power filters.it not only eliminate the harmonics, also it treat all type of voltage and current fluxuations and compensate the reactive power in distribution system. In this paper unified power quality conditioner with different control strategy is introduced to rectify the power quality issues and increase the efficiency of power quality. UPQC concern feedback system with PI controller is used to improve the performance of UPQC and compare the different strategy using MAT LAB/SIMULINK.

KEYWORDS: Power Quality, APF (active power filters)

I.INTRODUCTION

In general the power system consists of generation transmission and distribution. Apart from that generation, the transmission and distribution have the major problem called power quality issues. The power quality issues are voltage sag, voltage swell, interruption, harmonics, flickers etc. now a days we frequently using the many sensitive loads such as computer, led television etc. Due to poor power quality such equipment may failure. To diagnose this problem and also to improve the power quality we have only solution, called unified power quality conditioner.

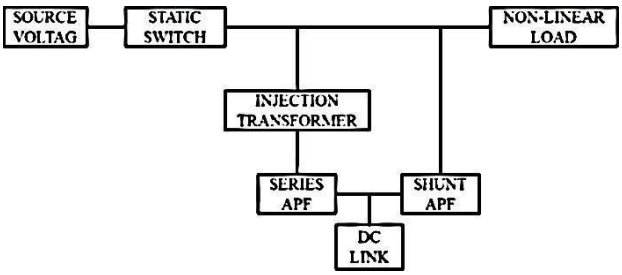


Fig: 1 Block Diagram of UPQC

From the block diagram the UPQC consist of two components named series active power filter and shunt active filter. It is also equalized with the DVR and D-STATCOM. The performance of voltage controlled voltage source inverter is acted as series active power filter, and the performance of current controlled voltage source inverter is acted as shunt active power filter. The both series active power filter and shunt active power filter are coupled with dc capacitor for dc link. The series APF is connected via an injection transformer with

the ac line. The isolation of voltage based distortion is done by the series APF and the isolation of current based problems is done by the shunt APF. Also it treats the reactive current of the load and improves the power of the system.

A.SERIES ACTIVE POWER FILTER

The series APF is a series element which can act as a controlled voltage source. It gives voltage of negative polarity harmonics by injection transformer. The basic circuits of series APF is shown in figure.

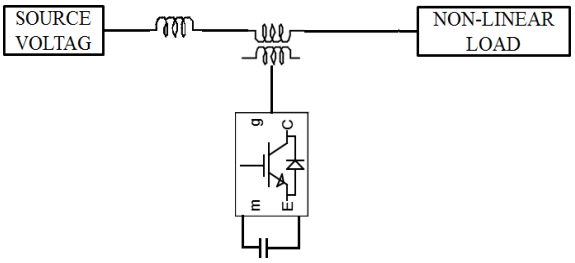


Fig: 2 Series Active Power Filter

The capacitor is energy storage wit self-supporting that is with reactive power exchange. If we use constant dc source then there exists only a real power exchange through voltage source inverter

B.SHUNT ACTIVE POWER FILTER

Shunt active power filter is a shunt connecting device which can be acts as controlled current source. It gives opposite current harmonics to clear current related problems. The performance of dc capacitor is same as the series APF. The basic circuit configuration shunt APF is shown in figure.

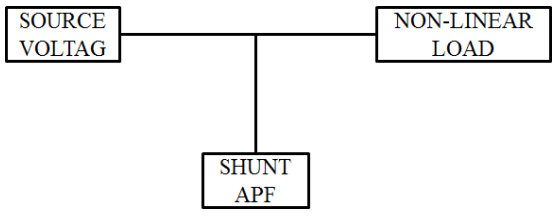


Fig: 03 Shunt Active Power Filter.

The function of shunt APF is dc link voltage regulation improvement of power factor by controlling reactive power.

II. DIFFERENT CONTROL STRATEGIES

In this paper there are three different control strategies are explained and compared by its simulation results. The various strategies are

1) A normal transmission line with source side and load side. To check its performance under 3 phase fault condition using mat lab/simulation. Here the line supply voltage is taken as 400v for each phase for injecting the harmonics we considered the RL load with the resistance value of 30Ω and the inductance value of $10e^{-3}$H
2) Transmission line with unified power quality conditioner is considered here the line supply voltage is take as 400v fir each phase for providing the harmonics we considered the RL load with the resistor value of 20Ω and inductor value of $50e^{-3}$H
3) Here the transmission line is operated through the UPQC with PI controller is introduced. The line supply voltage is 400v and resistor is considered as the non-linear load valued300Ω.

III. DESIGN OF CONTROLLERS

The use of power electronic controller in the electric power supply system has become very common. The controllers are used to improve the performance of the system with the help of feedback. The controller is applicable only for the closed loop system.

A.OPEN LOOP SYSTEM

The function of any electronic system is to automatically regulate the output and keep it within the systems desired input value or fixed point. If the

system desired input value or any other reason, the output of the system must respond accordingly and change itself to reflect the new input value. Likewise, if something happens to disturb the systems output without any change to the input value, the output responds and returning back to its previous set value. The electronic system was basically controlled only manually called open loop control system.

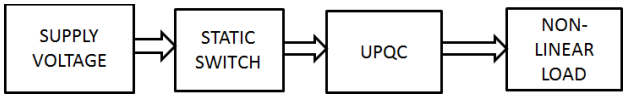


Fig: 04 Block Diagram of Open Loop System

B.CLOSED LOOP SYSTEM

The systems in which the output quantity has no effect upon the input to the control process are called open loop control systems, and the open loop systems are just that named non feedback systems. But the goal of any power system control is to measure, monitor, and control the process. One way in which we can obviously control the process is by monitoring its output and ‒feeding| some of it back to compare the actual output with the desired output so as to reduce the error, if the system disturbed then the output of the system back to the original or desired response. The measure of the output is known as feedback signal and the type of control system which uses feedback signals to control itself is called a Close-loop System.

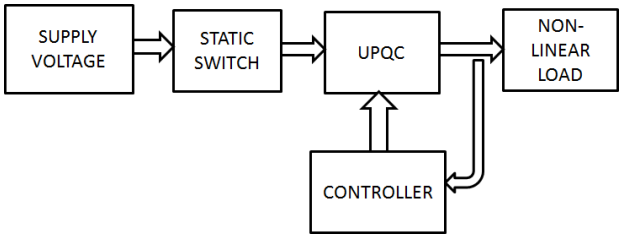


Fig: 05 Block Diagram of Closed Loop System

C.PI CONTROLLER

A PI controller calculates an error value as the difference between a measured process variable and desired set point. The controller attempts to reduce the error by adjusting the process control inputs. V_{dc} is sensed and compared with its reference V_{dc}*. Error signal is processed in a PI controller. The output of the PI controller is expressed as $I_{sp(n)}$. The output of controller has a limit ensures that the source provisions active power of the load and dc bus of the UPQC .A self-supported dc link of the UPQC is supplied by the active power. Thus, the dc voltage of the UPQC has a proper current.

IV. PERFORMANCE ANALYSIS OF UPQC

A. TRANSISSION LINE WITH POOR POWER QUALITY

In this paper first we are going to analyse the performance of transmission system without unified power quality conditioner. As the resulting leads the severe voltage distortion and losses. The mat lab simulation is analysed by following.

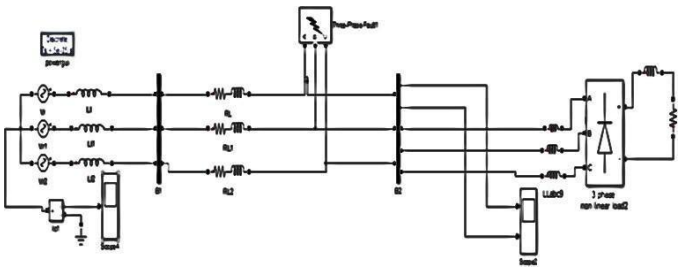


Fig 7: Transmission Line

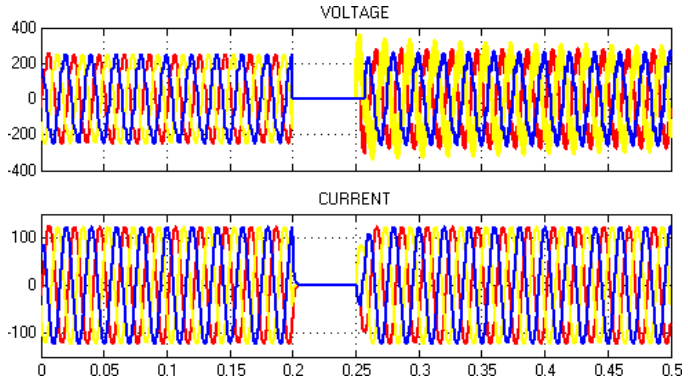


Fig 8: Output waveforms for transmission line

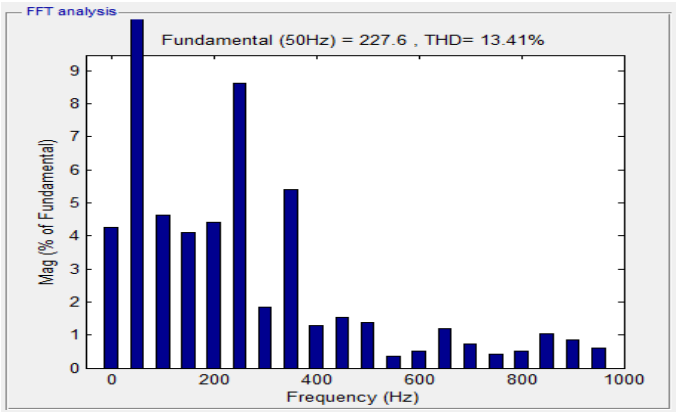


Fig 9: FFT Analysis for the transmission line

B.TRANSMISSION LINE WITH UPQC

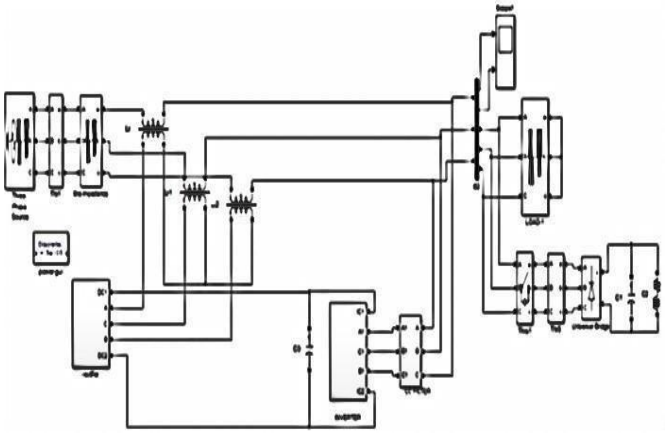


Fig 10: Transmission Line with UPQC

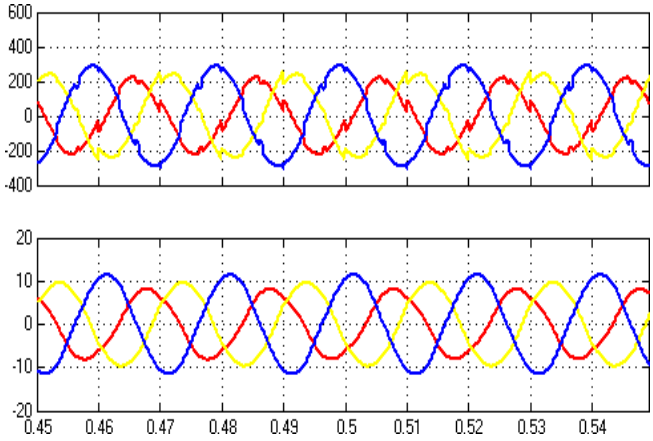


Fig 11: Output waveforms for with UPQC

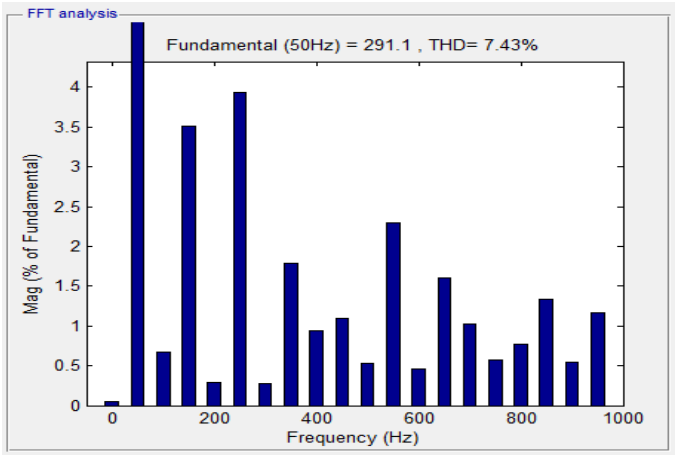


Fig 12: FFT Analysis for with UPQC

C. UPQC WITH PI CONTROLLER

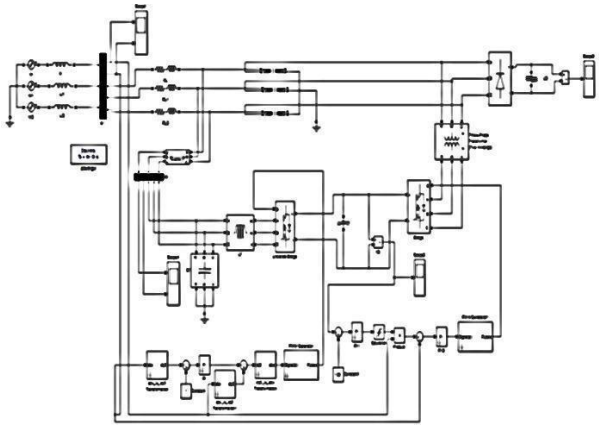


Fig 13: UPQC with PI Controller

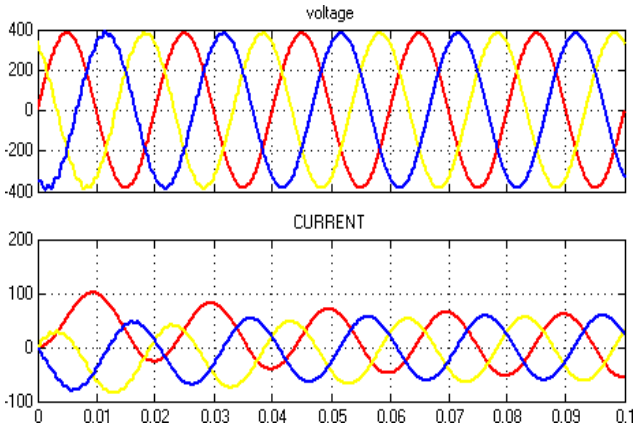


Fig 14: Output Waveforms for UPQC with PI

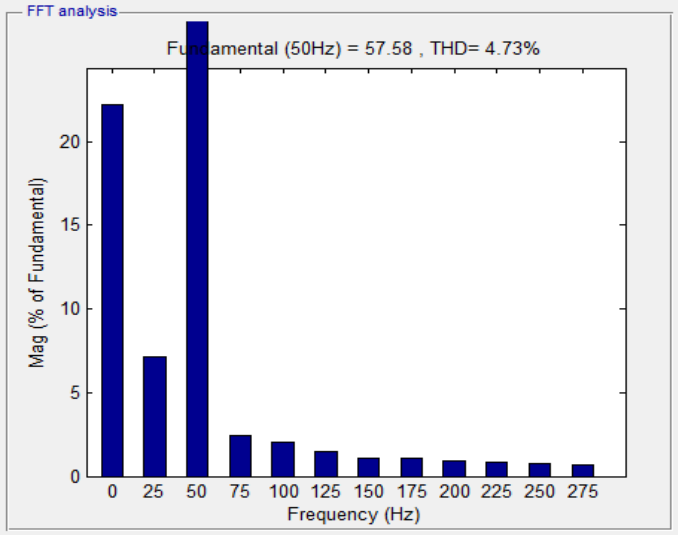


Fig 15: FFT Analysis for UPQC with PI

CONCLUSION

The different control strategy of UPQC was described and compared its performance using simulation. The power quality issues are almost reduced. The closed loop control schemes of current control, for the proposed UPQC have been investigated. Total harmonic distortion was analysed and that describes that the UPQC with PI controller provides more efficiency than the other strategies.

REFERENCES

[1] Vadirajacharya G. Kinhal, Promod Agarwal, and Hari Oam Gupta, ―Performance Investigation of Neural-Network-Based Unified Power-Quality Conditioner‖ IEEE Trans. On Power Delivery, Vol. 26, No. 1, January 2011

[2] E. W. Gunther and H. Mehta, ―A survey of distribution system power quality, ‖ IEEE Trans. Power Del., vol. 10, no. 1, pp. 322–329, January. 1995.

[3] Y. Chen, L. F. Sanchez, K. M. Smedley, and G. Chen, ―One-cycle controlled unified power quality conditioner for load side voltage sag compensation,‖ in Proc. Power Electronics Specialists Conf., 2005, vol. 36, pp. 282–288.

[4] R. E. King, Computational Intelligence in Control Engineering, ser. Control Eng. Basel, New York: Marcel Dekker.

[5] A. Zouidi, F. Fnaiech, and K. AL-Haddad, ―Neural network controlled three-phase three-wire shunt active power.

[6] W. M. Grady, M. J. Simony, and A. A. Noyola, ―Survey of active power line conditioning methodologies,‖ IEEE Trans. Power Del., vol. 5, no. 3, pp. 1536–1542, July. 1990.

[7] F. Kamron, ―Combined dead beat control of series—Parallel converter combination used as a universal power filter,‖ in Proc. IEEE Power Electronics Specialist Conf., 1995, pp. 196–201.

[8] H. Fujita and H. Akagi, ―The unified power quality conditioner: The integration of series active filter and shunt active filters,‖ in Proc. IEEE/ Power Eng. Soc. Power Electronics Specialist Conf., June. 1996, pp. 491–501.

[9] V. S. C. Raviraj and P. C. Sen, ―Comparative study of proportional integral, sliding mode and fuzzy logic controllers for power converters,‖ IEEE Trans. Ind. Appl., vol. 33, no. 2, pp. 518–524, March./April. 1997.

[10] J. H. Marks and T. C. Green, ―Predictive control of active filters,‖ in Proc. Inst. Elect. Eng. Conf. Power Electronics and Variable Speed Drives, September. 2000, pp. 18–23. filter,‖ in Proc. IEEE ISIE, Montreal, QC, Canada, July. 9–12, 2006, pp. 5–10

[11] R. El Shatshat, M. M. A. Salama, and M. Kazerani, ―Artificial intelligent controller for current source converter-based modular active power filters,‖ IEEE Trans. Power Del., vol. 19, no. 3, pp. 1314–1320, July. 2004.

[12] J. R. Vazquez and P. R. Salmer6n, ―Three-phase active power filter control using neural networks,‖ in Proc. 10th Mediterranean Electro Technical Conf., 2000, vol. III, pp. 924–927.

ISSN: 2454-9924

Hide me and Authenticate Implementation of Multi party Key Authentication and SAPA Protocol for Secured Data Transaction in Cloud

R. Archana #1, S. Prasanna *1

Mailam Engineering College, Mailam #1, *1

archanaravanan@gmail.com

Abstract - In the Cloud computing is an evolving data communicative design to determine the user data remotely stored in an online cloud server. Security solutions mainly focus on the authentication cannot be illegally accessed, but neglect a subtle privacy issue. In the proposed mechanism, there will be three Entities Users, Cloud Server & Trusted Third Party (TPA). Data Users are both Data Owners & Data Users. Every User will be registering with the Cloud Server. Cloud will be generating Pair wise Keys, Primary & Secondary Keys for both Cloud Server & Data User. Users 1 wants to Access the data of Users 2 then Keys are Shared Keys are generated and accordingly the Data is authorized for Usage. In our modified process, an Access key is generated while Registration with Cloud. After that only Shared Keys are generated. Finally a Mutual Access key is generated by the data owner to the data user and sent via Email. Data User will have to hide that Mutual Key in an Image called Steganography and sent to the Data Owner. Data is accessed by only after Verifying Mutual Key using Destaganography.

Keywords – Cloud, Trusted Third Party, Shared keys, Mutual key

1 Introduction

Cloud computing is a promising information technology architecture for both enterprises and individuals. It launches an attractive data storage and interactive paradigm with obvious advantages, including on-demand self-services, ubiquitous network access, and location independent resource pooling. Towards the cloud computing, typical service architecture is anything as a service (XaaS), in which infrastructures, platform, software, and others are applied for ubiquitous interconnections. Recent studies have been worked to promote the cloud computing evolve towards the internet of services. Subsequently, security and privacy issues are becoming key concerns with the increasing popularity of cloud services. Conventional security approaches mainly focus on the strong authentication to realize that a user can remotely access its own data in on demand mode. Along with the diversity of the application requirements, users may want to access and share each other's authorized data fields to achieve productive benefits, which brings new security and privacy challenges for the cloud storage. An example is introduced to identify the main motivation. In the cloud storage based supply chain management, there are various interest groups (e.g., supplier, carrier, and retailer) in the system. Each group owns its users which are permitted to access the authorized data

fields, and different users own relatively independent access authorities. It means that any two users from diverse groups should access different data fields of the same file. There into, a supplier may want to access a carrier's data fields, but it is not sure whether the carrier will allow its access request. If the carrier refuses its request, the supplier's access desire will be revealed along with nothing obtained towards the desired data fields. Actually, the supplier may not send the access request or withdraw the unaccepted request in advance if it firmly knows that its request will be refused by the carrier. It is unreasonable to thoroughly disclose the supplier's private information without any privacy considerations. In the cloud environments, a reasonable security protocol should achieve the following requirements. 1) Authentication: a legal user can access its own data fields, only the authorized partial or entire data fields can be identified by the legal user, and any forged or tampered data fields cannot deceive the legal user. 2) Data anonymity: any irrelevant entity cannot recognize the exchanged data and communication state even it intercepts the exchanged messages via an open channel. 3) User privacy: any irrelevant entity cannot know or guess a user's access desire, which represents a user's interest in another user's authorized data fields. If and only if the both users have mutual interests in each other's authorized data fields, the cloud server will inform the two users to realize the access permission sharing. 4) Forward security: any adversary cannot correlate two communication sessions to derive the prior interrogations according to the currently captured messages. In this paper, we address the aforementioned privacy issue to propose a shared authority based privacy-preserving authentication protocol (SAPA) for the cloud data storage, which realizes authentication and authorization without compromising a user's private information. The main contributions are as follows. 1) Identify a new privacy challenge in cloud storage, and address a subtle privacy issue during a user challenging the cloud server for data sharing, in which the challenged request itself cannot reveal the user's privacy no matter whether or not it can obtain the access authority. 2) Propose an authentication protocol to enhance a user's access request related privacy, and the shared access authority is achieved by anonymous access request matching mechanism. 3) Apply cipher text-policy attribute based access control to realize that a user can reliably access its own data fields, and adopt the proxy re-encryption to provide temp authorized data sharing among multiple users.

2 Related Work

In the existing mechanism, security solutions mainly focus on the authentication cannot be illegally accessed, but neglect a subtle privacy issue. Here it provides some drawbacks are, Less Security, Data hacking, missing privacy

2.1 Proposed Mechanism

In the proposed mechanism, there will be three Entities Users, Cloud Server & Trusted Third Party (TPA). Data Users are both Data

Owners & Data Users. Every User will be registering with the Cloud Server. Cloud will be generating Pair wise Keys, Primary & Secondary Keys for both Cloud Server & Data User. Users 1 wants to Access the data of Users 2 then Keys are Shared Keys are generated and accordingly the Data is authorized for Usage. In our modified mechanism, an Access key is generated while Registration with Cloud. After that only Shared Keys are generated. Finally a Mutual Access key is generated by the data owner to the data user and sent via Email. Data User will have to hide that Mutual Key in an Image called Steganography and sent to the Data Owner. Data is accessed by only after Verifying Mutual Key using Destaganography. Here it provides some benefits are High security, Data integrity, easily find the attacker.

3 System Design

System model for the cloud storage architecture, which includes three main network entities: users (U x), a cloud server (S), and a trusted third party. An individual or group entity, which owns its data stored in the cloud for online data storage and computing. Different users may be affiliated with a common organization, and are assigned with independent authorities on certain data fields. Cloud server an entity, which is managed by a particular cloud service provider or cloud application operator to provide data storage and computing services. The cloud server is regarded as an entity with unrestricted storage and computational resources. Trusted third party, an optional and neutral entity, which has advanced capabilities on behalf of the users, to perform data public auditing and dispute arbitration. In the cloud storage, a user remotely stores its data via online infrastructures, flat forms, or software for cloud services, which are operated in the distributed, parallel, and cooperative modes. During cloud data accessing, the user autonomously interacts with the cloud server without external interferences, and is assigned with the full and independent authority on its own data fields. It is necessary to guarantee that the users' outsourced data cannot be unauthorized accessed by other users, and is of critical importance to ensure the private information during the users' data access challenges. In some scenarios, there are multiple users in a system (e.g., supply chain management), and the users could have different affiliation attributes from different interest groups. One of the users may want to access other associated users' data fields to achieve bi-directional data sharing, but it cares about two aspects: whether the aimed user would like to share its data fields, and how to avoid exposing its access request if the aimed user declines or ignores its challenge. In the paper, we pay more attention on the process of data access control and access authority sharing other than the specific file oriented cloud data management. In the system model, assume that point-to-point communication channels between users and a cloud server are reliable with the protection of secure shell protocol (SSH). The related authentication handshakes are not highlighted in the following protocol presentation. Towards the trust model, there

are no full trust relationships between a cloud server S and a user U x S is semi-honest and curious. Being semi-honest means that S can be regarded as an entity that appropriately follows the protocol procedure. Being curious means that S may attempt to obtain U's private information (e.g., data content, and user preferences). It means that S is under the supervision of its cloud provider or operator, but may be interested in viewing users' privacy. In the passive or honest-but curious model, S cannot tamper with the users' data to maintain the system normal operation with undetected monitoring x.

4 Literature review

1. **P. Mell and T. Grance, Draft NIST Working Definition of Cloud Computing," Nat'**

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

2. **Moreno-Vozmediano, R., Key Challenges in Cloud Computing: Enabling the Future Internet of Services**

Cloud computing will play a major role in the future Internet of Services, enabling on-demand provisioning of applications, platforms, and computing infrastructures. However, the cloud community must address several technology challenges to turn this vision into reality. Specific issues relate to deploying future infrastructure-as-a-service clouds and include efficiently managing such clouds to deliver scalable and elastic service platforms on demand, developing cloud aggregation architectures and technologies that let cloud providers collaborate and interoperate, and improving cloud infrastructures' security, reliability, and energy efficiency.

3. **Kai Hwang, Trusted Cloud Computing with Secure Resources and Data Coloring**

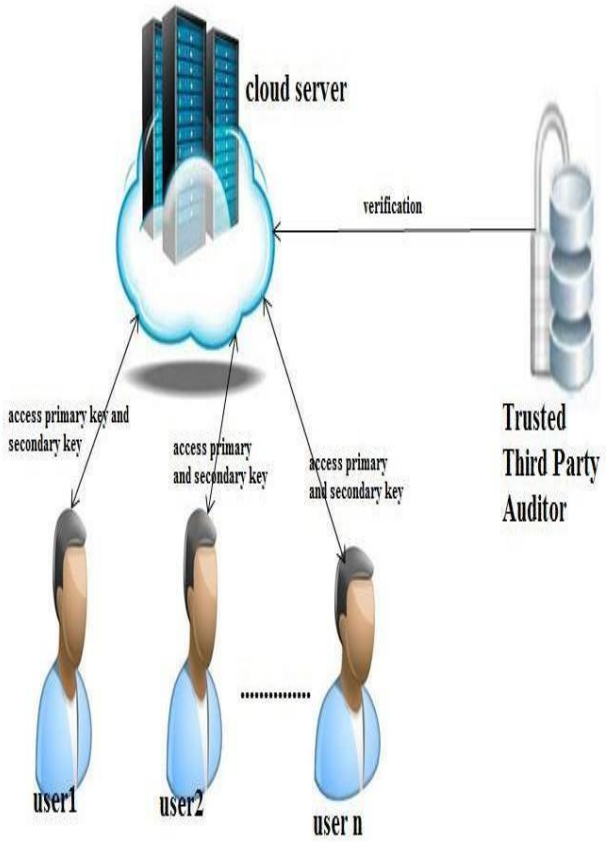
Trust and security have prevented businesses from fully accepting cloud platforms. To protect clouds, providers must first secure virtualized data center resources, uphold user privacy, and preserve data integrity. The authors suggest using a trust-overlay network over multiple data centers to implement a reputation system for establishing trust between service providers and data owners. Data coloring and software watermarking techniques protect shared data objects and massively distributed software modules. These techniques safeguard multi-way authentications, enable single sign-on in the cloud, and tighten access control for sensitive data in both public and private clouds.

4. **Jianyong Chen, On-Demand Security Architecture for Cloud Computing**

An architecture thatdifferentiates security ac cording to service-specific characteristics avoids an unnecessary drain on IT resources by protecting a variety of cloud computing services at just the right level.

5 Architecture Design

In the system model, assume that point-to-point communication channels between users and a cloud server are reliable with the protection of secure shell protocol (SSH). The related authentication handshakes are not highlighted in the following protocol presentation. Towards the trust model, there are no full trust relationships between a cloud server S and a user Ux. S is semi-honest and curious. Being semi-honest means that S can be regarded as an entity that appropriately follows the protocol procedure.



6 Conclusion

From this hide me and Authenticate Implementation of Multi party Key Authentication and SAPA Protocol for Secured Data Transaction in Cloud have been implemented, In this work, we have identified a new privacy challenge during data accessing in the cloud computing to achieve privacy-preserving access authority sharing. Authentication is established to guarantee data confidentiality and data integrity. Data anonymity is achieved since the wrapped values are exchanged during transmission. User privacy is enhanced by anonymous access requests to privately

inform the cloud server about the users' access desires. Forward security is realized by the session identifiers to prevent the session correlation. It indicates that the proposed scheme is possibly applied for privacy preservation in cloud applications. In future, we also demonstrate the efficient system performance.

7 References

[1] A. Barsoum and A. Hasan, "Enabling Dynamic Data and Indirect Mutual Trust for Cloud Computing Storage Systems, IEEE Trans. Parallel and Distributed Systems, vol. 24, no. 12, pp. 2375-2385, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6392165, Dec. 2013.

[2] A. Mishra, R. Jain, and A. Durresi, "Cloud Computing: Networking and Communication Challenges," IEEE Comm. Magazine, vol. 50, no. 9, pp. 24-25, Sept. 2012.

[3] C. Wang, K. Ren, W. Lou, and J. Li, "Toward Publicly Auditable Secure Cloud Data Storage Services," IEEE Network, vol. 24, no. 4, pp. 19-24, July/Aug. 2010.

[4] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, "Toward Secure and Dependable Storage Services in Cloud Computing," IEEE Trans. Services Computing, vol. 5, no. 2, pp. 220-232, Apr.-June 2012.

[5] H. Wang, "Proxy Provable Data Possession in Public Clouds," IEEE Trans. Services Computing, vol. 6, no. 4, pp. 551-559,http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6357181,Oct.-Dec. 2012.

[6] H. Zhuo, S. Zhong, and N. Yu, "A Privacy-Preserving Remote Data Integrity Checking Protocol with Data Dynamics and Public Verifiability," IEEE Trans. Knowledge and Data Eng., vol. 23, no. 9, pp. 1432-1437, Sept. 2011.

[7] J. Chen, Y. Wang, and X. Wang, "On-Demand Security Architecture for Cloud Computing," Computer, vol. 45, no. 7, pp. 73-78, 2012.

[8] K. Hwang and D. Li, "Trusted Cloud Computing with Secure Resources and Data Coloring," IEEE Internet Computing, vol. 14, no. 5, pp. 14-22, Sept./Oct. 2010.

[9] K. Yang and X. Jia, "An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing," IEEE Trans. Parallel and Distributed Systems, vol. 24, no. 9, pp. 1717-1726, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6311398, Sept. 2013.

[10] L.A. Dunning and R. Kresman, "Privacy Preserving Data Sharing with Anonymous ID Assignment," IEEE Trans. Information Forensics and Security, vol. 8, no. 2, pp. 402-413, Feb. 2013.

[11] P. Mell and T. Grance, "Draft NIST Working Definition of Cloud Computing," Nat'l Inst. of Standards and Technology, 2009.

[12] R. Canetti, "Universally Composable Security: A New Paradigm for Cryptographic Protocols," Proc. 42nd IEEE Symp. Foundations of Computer Science (FOCS '01), pp. 136-145, Oct. 2001.

[13] R. Moreno-Vozmediano, R.S. Montero, and I.M. Llorente, "Key Challenges in Cloud Computing to Enable the Future Internet of Services,"IEEEInternet Computing,vol.17,no. 4,pp. 18-25,http:// ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6203493, July/Aug.2013.

[14] R. Sanchez, F. Almenares, P. Arias, D. Dıaz-Sanchez, and A.Marın, "Enhancing Privacy and Dynamic Federation in IdM for Consumer Cloud Computing," IEEE Trans. Consumer Electronics, vol. 58, no. 1, pp. 95-103, Feb. 2012.

[15] S. Ruj, M. Stojmenovic, and A. Nayak, "Decentralized Access Control with Anonymous Authentication for Securing Data in Clouds," IEEE Trans. Parallel and Distributed Systems, vol. 25, no. 2, pp. 384-394,http://ieeexplore.ieee.org/stamp/stamp.jsp? tp=&arnumber=6463404, Feb. 2014.

[16] S. Sundareswaran, A.C. Squicciarini, and D. Lin, "Ensuring Distributed Accountability for Data Sharing in the Cloud," IEEE Trans. Dependable and Secure Computing, vol. 9, no. 4, pp. 556-568, July/Aug. 2012.

[17] X. Liu, Y. Zhang, B. Wang, and J. Yan, "Mona: Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud," IEEE Trans. Parallel and Distributed Systems, vol. 24, no. 6, pp. 1182-1191, http:// ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6374615, June 2013.

[18] Y. Tang, P.C. Lee, J.C.S. Lui, and R. Perlman, "Secure Overlay Cloud Storage with Access Control and Assured Deletion," IEEE Trans. Dependable and Secure Computing, vol. 9, no. 6, pp. 903-916, Nov./Dec. 2012.

[19] Y. Xiao, C. Lin, Y. Jiang, X. Chu, and F. Liu, "An Efficient Privacy-Preserving Publish-Subscribe Service Scheme for Cloud Computing," Proc. IEEE GLOBECOM '10, Dec.2010.

[20] Y. Zhu, H. Hu, G. Ahn, and M. Yu, "Cooperative Provable Data Possession for Integrity Verification in Multi-Cloud Storage," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 12, pp. 2231-2244, Dec. 2012.

ISSN: 2454-9924

Mitigating Web Spam Taxonomy for Mobile App using Link Pruning and Reweighting Algorithm

V. Suriya #1, R. Mohan *1

Mailam Engineering College, Mailam #1 *1

Suriya.mangai@gmail.com #1

Abstract - In mobile application the fraud activities are widely spread and produce data loss. Ranking fraud in the mobile App market refers to fraudulent or deceptive activities which have a purpose of crash up the Apps in the popularity list. In real, it becomes more and more frequent for App developers to use covered means, such as inflating their Apps' sales or posting phony App ratings, to commit ranking fraud. In existing methods the prevention of fraud activities are not clearly determined ever. In this proposed system, we provide a rounded view of ranking fraud and web spam to detect and reduce the ranking fraud in mobile Apps systems. In this project, we provides link cutting and reweighting algorithm to find the web spam analysis and provide a broad coverage of various web spam forms. Using above algorithm Link spam and Click spam are determined. Link spam Adding links that point to the spammer's web site increases the page rankings for the site in the App Store. Similarly click spam, clicking ad banners without any motivation of purchasing the product.

Keywords – Mobile application, Fraud, Prevention, Spam, Rating

1 Introduction

The number of mobile Apps has grown at a breathtaking rate over the past few years. For example, as of the end of April 2013, there are more than 1.6 million Apps at Apple's App store and Google Play. To stimulate the development of mobile Apps, many App stores launched daily App leaderboards, which demonstrate the chart rankings of most popular Apps. Indeed, the App leaderboard is one of the most

ISSN: 2454-9924

important ways for promoting mobile Apps. A higher rank on the leaderboard usually leads to a huge number of downloads and million dollars in revenue. Therefore, App developers tend to explore various ways such as advertising campaigns to promote their Apps in order to have their Apps ranked as high as possible in such App leaderboards. However, as a recent trend, instead of relying on traditional marketing solutions, shady App developers resort to

ISSN: 2454-9924

some fraudulent means to deliberately boost their Apps and eventually manipulate the chart rankings on an App store. This is usually implemented by "Spam" inflate the App downloads, ratings and reviews in a very short time. Spam pervades any information system, be it e-mail or web, social, blog or reviews platform. In the literature, while there is some related work, such as web ranking spam detection, online review spam detection, and mobile App recommendation the problem of detecting ranking fraud for mobile Apps is still underexplored. To fill this crucial void, in this project, we propose to develop a ranking fraud detection system for mobile Apps. Generally speaking, web spam manifests itself as a web content generated deliberately for the purpose of triggering unjustifiably favorable relevance or importance of some web page or pages. It is worth mentioning that the necessity of dealing with the malicious content in a corpus is a key distinctive feature of adversarial information retrieval in comparison with the traditional information retrieval, where algorithms operate on a clean benchmark data set or in an intranet of a corporation. Daily App leaderboard became a de facto place to start information acquisition of mobile App.

ISSN: 2454-9924

Though due to web spam phenomenon, search results are not always as good as desired. Moreover, spam evolves that makes the problem of providing high quality search even more challenging. Over the last decade research on adversarial information retrieval has gained a lot of interest both from academia and industry. In this project we present a holistic view of web spam detection for mobile App. When apps are downloaded from app store, Link spam and Click spam are detected using link pruning and reweighting algorithm. Link spam and click spam are comes under web spam. During app download the rating will be automatically incremented if the app is not affected by any spam (link and click spam), otherwise the rating standard with previous one. A spammer creates a page which looks absolutely innocent and may be even authoritative (though it is much more expensive), but links to the spammer's target pages. In this case an organically aggregated PageRank (authority) score is propagated further to target pages and allows them to be ranked higher. More aggressive form of a link spam schema is hijacking, when spammers first hack a reputable website and then use it as a part of their link farm. Spammers can also collude by participating

ISSN: 2454-9924

in link exchange schemes in order to achieve higher scale, higher in-link counts, or other goals. We also consider redirection as an instant type of link spam. Here the spamming scheme works as follows. First, a link spam page achieves high ranking in a user review page by boosting techniques. But when the page is requested by a user, they don't actually see it; they get redirected to a target page. There are various ways to achieve redirection. The easiest approach is to set a page refresh time to zero and initialize a refresh URL attribute with a URL of a target page. More sophisticated approach is to use page level scripts that aren't usually executed by crawlers and hence more effective from spammers point of view. Since web use click stream data as an implicit feedback to tune ranking functions, spammers are eager to generate fraudulent clicks with the intention to bias those functions towards their websites. To achieve this goal spammers submit queries to a search engine and then click on links pointing to their target pages. To hide anomalous behavior they deploy click scripts on multiple machines or even in large botnets. The other incentive of spammers to generate fraudulent clicks comes from online advertising. In this case, in reverse,

ISSN: 2454-9924

spammers click on ads of competitors in order to decrease their budgets, make them zero, and place the ads on the same spot. The App leaderboard is one of the most important ways for promoting mobile Apps. A higher rank App on the leaderboard usually leads to a huge number of downloads and million dollars in revenue. So, some fraudulent had happen to boost their Apps and eventually manipulate the chart rankings on an App store. In this project, we developed a ranking fraud detection system for mobile Apps. When apps are downloaded from app store, "Spam" inflate the App downloads, ratings and reviews in a very short time. Spam pervades any information system, be it e-mail or web, social, blog or reviews platform. Different types of spams are available in web. In this, we detect and overcome "Link spam" and "Click spam" using link pruning and reweighting algorithm.

2 Related Work

App stores launched daily App leaderboards to stimulate the development of mobile Apps. The App leaderboard is one of the most important ways for promoting mobile Apps. A higher rank App on the

ISSN: 2454-9924

leaderboard usually leads to a huge number of downloads. Therefore, App developers advertising campaigns to promote their Apps in order to have their Apps ranked as high as possible in such App leaderboards. Instead of relying on traditional marketing solutions shady App developers do some fraudulent using bot farms to boost their Apps chart ranking on App store. Web ranking spam detection, online review spam detection, and mobile App recommendation are still under explored. Moreover, spam evolves that makes the problem of providing high quality search even more challenging. Here it provides some drawbacks are, It is difficult to detect when fraud happens, It is difficult to manually label ranking fraud for each App, Is not easy to identify and confirm ranking fraud, Search results are not always good, Problem of providing high quality search, Web spam is not detect.

2.1 Proposed work

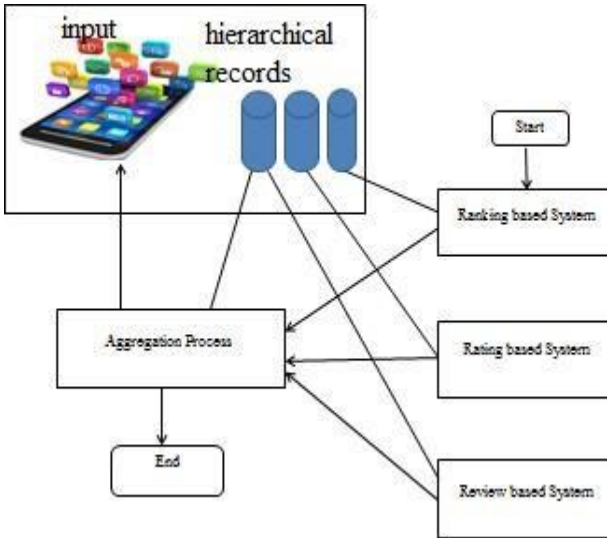
Apple's App store and Google Play became a de facto place to search and download Mobile Apps Store. Though due to web spam phenomenon, search results are not always as good as desired. So, we had to detect and avoid the web spam taxonomy. In our proposed work, we present a systematic

ISSN: 2454-9924

review of web spam detection techniques with the focus on algorithms and underlying principles. Link spam and Click spam both are web spam discussed in our proposed work. Link spam Adding links that point to the spammer's web site increases the page rankings for the site in the App Store. Similarly click spam, clicking ad banners without any intention of purchasing the product. Clicking the ads countless times can make dishonest rankings in Mobile App Store. Link pruning and reweighting algorithms are used here to detect and avoid the web spam. Link pruning and reweighting algorithms detect the “nepotistic links”, links that present for reasons rather than merit, for instance, navigational links on a website or links between pages in a link farm and also reports its resistance to fraudulent clicks. Here, it provides some benefits are, Automatic rating, Web spam detected, Link spam and click spam are detected, providing high quality search result.

3 Architecture

ISSN: 2454-9924



3.1 Ranking based System

A leading session is composed of several leading events. There-fore, we should first analyze the basic characteristics of leading events for extracting fraud evidences. By analyzing the Apps' historical ranking records, we observe that Apps' ranking behaviors in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase. Specifically, in each leading event, an App's ranking first increases to a peak position in the leaderboard (i.e., rising phase).

3.2 Rating based system

The ranking based evidences are useful for ranking fraud detection. However, sometimes, it is not sufficient to only use ranking based evidences. For example, some Apps created by the famous developers, such as Game loft, may have some leading events with large values of u due to the developers' credibility and the "word-of-mouth" advertising effect. Moreover, some of the legal marketing services, such as "limited-time discount", may also result in significant ranking based evidences. To solve this issue, we also study how to extract fraud evidences from Apps' historical rating records.

3.3 Review Based System

Besides ratings, most of the App stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the most important perspectives of App ranking fraud. Specifically, before downloading or purchasing a new mobile App, users often first read its historical reviews to ease their decision making, and a mobile App contains more positive reviews may attract more

ISSN: 2454-9924

users to download.

ISSN: 2454-9924

4 Methodology

4.1 Link Pruning and Reweighting

Notices that PageRank score of pages that achieved high ranks by link-spamming techniques correlates with the damping factor c. Using this observation authors identify suspicious nodes, whose correlation is higher than a threshold and down weight outgoing links for them with some function proportional to correlation. They also prove that spammers can amplify PageRank score by at most 1c and experimentally show that even two-node collusion can yield a big PageRank amplification. Where they show that due to the power law distribution of PageRank, the increase in PageRank is negligible for top-ranked pages.

4.2 Detection of Spam

Interesting idea to prevent click spam is proposed personalized ranking functions, as being more robust, to prevent click fraud manipulation. We present a utility-based framework allowing judging when it is economically reasonable to hire spammers to promote a website. The performs experimental study demonstrating that personalized ranking is resistant to spammers manipulations and diminishes financial incentives of site owners to hire spammers. The work studies the robustness of the standard click-through-based ranking function construction process and also reports its resistance to fraudulent clicks.

4.3 Web Graph

We model the Web as a graph with vertices, representing web pages, and directed weighted edges, representing hyperlinks between pages. If a web page(pi) has multiple hyperlinks to a page(pj), we will collapse all these links into one edge. Self-loops aren't allowed. We denote a set of pages linked by a page pi as Out(pi) and a set of pages pointing to pi as In(pi). Finally, each edge can have an associated non-negative weight.

4.4 Page Ranking Design

PageRank uses link information to compute global importance scores for all pages on the web. The key underlying idea is that a link from a page pi to a page pj shows an endorsement or trust of page pi in page pj, and the algorithm follows the repeated improvement principle. The true

ISSN: 2454-9924

score is computed as a convergence point of

ISSN: 2454-9924

an iterative updating process. The most popular and simple way to introduce PageRank is a linear system formulation.

5 Conclusion

From this, Detection of fraud in Mobile Application using Ranking has been implemented. In this project, we developed a web spam detection system for mobile Apps. To draw a general picture of the web spam phenomenon, we first provide numeric estimates of spam on the Web, discuss how spam affects users rating for mobile apps, and motivate academic research. In our project, we present a systematic review of web spam detection techniques with the focus on algorithms and underlying principles. Link spam and Click spam both are web spam discussed in our work. According to this work, web spam detection research has gone through a few generations: starting from simple content based methods to approaches using sophisticated link mining and user behaviour mining techniques.

6 References

[1] (2014). [Online]. Available: http://en.wikipedia.org/wiki/ cohen's_kappa

[2] (2014). [Online]. Available: http://en.wikipedia.org/wiki/information_retrieval

[3] (2012). [Online]. Available: https://developer.apple.com/news/index.php?id=02062012a

[4] (2012). [Online]. Available: http://venturebeat.com/2012/07/03/apples-crackdown-on-app-ranking-manipulation/

[5] (2012). [Online]. Available: http://www.ibtimes.com/applethreatens-crackdown-biggest-app-store-ranking-fraud-406764

[6] (2012). [Online]. Available: http://www.lextek.com/manuals/onix/index.html

[7] (2012). [Online]. Available: http://www.ling.gu.se/lager/mogul/porter-stemmer.

[8] A. Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach. Learn., 2007, pp. 616–623.

[9] A. Klementiev, D. Roth, and K. Small, "Unsupervised rank aggregation with

ISSN: 2454-9924

distance-based models," in Proc. 25th Int. Conf. Mach.Learn., 2008, pp. 472–479.

[10] A. Klementiev, D. Roth, K. Small, and I. Titov, "Unsupervised rank aggregation with domain-specific expertise," in Proc. 21st Int. Joint Conf. Artif. Intell., 2009, pp. 1101–1106.

[11] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 632–640.

[12] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in Proc. 15th Int. Conf. World Wide Web, 2006, pp. 83–92.

[13] D. F. Gleich and L.-h. Lim, "Rank aggregation via nuclear norm minimization," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011, pp. 60–68.

[14] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., pp. 993–1022, 2003.

[15] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product

ISSN: 2454-9924

review spammers using rating behaviors," in Proc. 19th ACMInt. Conf. Inform. Knowl. Manage., 2010, pp. 939–948.

[16] G. Heinrich, Parameter estimation for text analysis, " Univ. Leipzig, Leipzig, Germany, Tech. Rep., http://faculty.cs.byu.edu/~ringger/CS601R/papers/Heinrich-GibbsLDA.pdf, 2008.

[17] J. Kivinen and M. K. Warmuth, "Additive versus exponentiated gradient updates for linear prediction," in Proc. 27th Annu. ACM Symp. Theory Comput., 1995, pp. 209–218.

[18] L. Azzopardi, M. Girolami, and K. V. Risjbergen, "Investigating the relationship between language model perplexity and ir precision-recall measures," in Proc. 26th Int. Conf. Res. Develop. Inform. Retrieval, 2003, pp. 369–370.

[19] N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 219–230.

[20] T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228–5235, 2004.

[21] Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, "A taxi driving fraud detection

ISSN: 2454-9924

system," in Proc. IEEE 11th Int. Conf. Data Mining, 2011, pp. 181–190.

[22] Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li, "Supervised rank aggregation," in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 481–490.

ISSN: 2454-9924

N Joint Authorities: Integration of Secured Cloud Data Access Privilege with Attribute based Encryption over Distributed Authorities

G. Saraswathi #1, J. Jayapriya *2

Mailam Engineering College, Mailam #1, *2

sarasgovin@gmail.com #1

Abstract—Cloud computing is a rapid growth field in computer technology, which provides flexible, on-demand, and low-cost usage of computing resources, but the data is deploy to some cloud providers, and variety privacy concerns emerge from it. Variety schemes based on the attribute-based encryption have been implemented to secure the cloud storage. Nevertheless, most work depends on the data contents privacy and the access control, while less interest is paid to the privilege control and the identity privacy. In this, we implement a semi nameless privilege control scheme nameless Control to address not only the data privacy, but also the user identity privacy in existing access control schemes. Nameless Control decentralizes the central authority to limit the identity leakage and thus achieves s. Besides, it also creates the file access control to the privilege control, by which privileges of all operations on the cloud data can be maintained in a fine-grained manner. Frequently, we provide the nameless Control-F, which fully determines the identity leakage and achieve the full anonymity. Finally, this proposed system provides, high performance efficiency and security in cloud Storage.

Keywords: Nameless, Privacy, Semi nameless, Authority, Encryption

1. Introduction

CLOUD computing is a revolutionary computing technique, by which computing resources are provided dynamically via Internet and the data storage and computation are outsourced to someone or some party in a _cloud‘. It greatly attracts attention and interest from both academia

ISSN: 2454-9924

and industry due to the profitability, but it also has at least three challenges that must be handled before coming to our real life to the best of our knowledge. First of all, data confidentiality should be guaranteed. The data privacy is not only about the data contents. Since the most attractive part of the cloud computing is the computation outsourcing, it is far beyond enough to just

ISSN: 2454-9924

conduct an access control. More likely, users want to control the privileges of data manipulation over other users or cloud servers. This is because when sensitive information or computation is outsourced to the cloud servers or another user, which is out of users‘ control in most cases, privacy risks would rise dramatically because the servers might illegally inspect users‘ data and access sensitive information, or other users might be able to infer sensitive information from the outsourced computation. Therefore, not only the access but also the operation should be controlled. Secondly, personal information (defined by each user‘s attributes set) is at risk because one‘s identity is authenticated based on his information for the purpose of access control (or privilege control in this paper). As people are becoming more concerned about their identity privacy these days, the identity privacy also needs to be protected before the cloud enters our life. Preferably, any authority or server alone should not know any client‘s personal information. Last but not least, the cloud computing system should be resilient in the case of security breach in which some part of the system is compromised by attackers. Various techniques have been proposed to protect the data contents privacy via access control. Identity-based encryption (IBE) was first introduced by Shamir, in which the sender of a message can specify an identity such that only a receiver with matching identity can decrypt it. Few years later, Fuzzy Identity-Based Encryption is proposed, which is also known as Attribute-Based Encryption (ABE). In such encryption

ISSN: 2454-9924

scheme, an identity is viewed as a set of descriptive attributes, and decryption is possible if a decrypt‘s identity has some overlaps with one specified in the cipher text. Soon after, more general tree-based ABE schemes, Key-Policy Attribute-Based Encryption (KP-ABE) and Cipher text-Policy Attribute- Based Encryption (CP-ABE), are presented to express more general condition than simple ‗overlap‘. They are counterparts to each other in the sense that the decision of encryption policy (who can or cannot decrypt the message) is made by different parties.

2. Related Work

In a multi-authority system is presented which each user has an ID and they can interact with each regenerator (authority) using different pseudonyms. One user

‗different pseudonyms are tied to his private key, but regenerators never know about the private keys, and thus there not able to link multiple pseudonyms belonging to the same user. Also, the whole attributes set is divided into N disjoint sets and managed by N attributes authorities. In this setting, each authority knows only a part of a user‘s attributes, which are not enough to figure out the user ‗identity considered the basicthreshold-based KP-ABE, which lacks generality in the encryption policy expression. Many attributed encryption schemes having multiple authorities have been proposed afterwards, but they either also employ threshold-based ABE, or have a semi-honest central authority, or cannot tolerate arbitrarily many users collusion

ISSN: 2454-9924

attack. The work is the most similar ones to ours in that they also tried to decentralize the central authority in the CP-ABE into multiple ones. Use a LSSS matrix as an access structure, but their scheme only converts the AND, OR gates to the LSSS matrix which limits their encryption policy to Boolean formula, while we inherit the flexibility of the access tree having thresholates. Muller et al. also supports only Disjunctive Normal Form (DNF) in their encryption policy. Besides the fact that we can express arbitrarily general encryption policy, our system also tolerates the compromise attack towards attributes authorities, which is not covered in many existing works. Recently, there also appeared traceable multi-authority ABE and, which are on the opposite direction of ours those schemes introduce accountability such that malicious users‘ keys can be traced. On the other hand, similar direction as ours can be found in, who try to hide encryption policy in the cipher texts, but their solutions do not prevent the attribute disclosure in the key generation phase. To some extent, these three works and ours complement each other the sense that the combination of these two types protection will lead to a completely anonymous ABE. A multi-authority system is presented in which each user has an ID and they can interact with each key generator (authority) using different pseudonyms. One user‘s different pseudonyms are tied to his private key, but key generators never know about the private keys, and thus they are not able to link multiple pseudonyms belonging to the same user. Also, the whole attributes set is divided

ISSN: 2454-9924

into N disjoint sets and managed by N attributes authorities. In this setting, each authority knows only a part of any user‘s attributes, which are not enough to figure out the user‘s identity. However, the schemeproposed by Chase et al. considered the basic threshold-based KP-ABE, which lacks generality in the encryption policy expression. Many attribute based encryption schemes having multiple authorities have been proposed afterwards, but they either also employ a threshold-based ABE, or have a semi-honest central authority, or cannottolerate arbitrarily many users‘ collusion attack are the most similar ones to ours in that they also tried to decentralize the centralauthority in the CP-ABE into multiple one, Use a LSSS matrix as an access structure, but their scheme only converts the AND, OR gates to the LSSS matrix, which limits their encryption policy to boo lean formula, while we inherit the flexibility of the access tree having threshold gates. Muller et al. also supports only Disjunctive Normal Form (DNF) in their encryption policy. Besides the fact that we can express arbitrarily general encryption policy, our system also tolerates the compromise attack towards attributes authorities, which is not covered in many existing works.

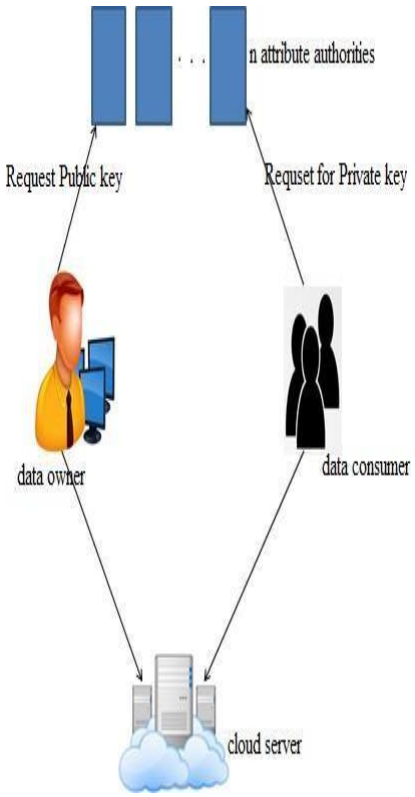
3. Proposed Work

Therefore, we propose Nameless Control and Nameless Control-F (Fig. 1) to allowcloud servers to control users‘ access privileges without knowing their identity information. Their main merits are: The proposed schemes are able to protect user‘s

ISSN: 2454-9924

privacy against each single authority. Partial information is disclosed in Nameless Control and no information is disclosed in Nameless Control-F. The proposed schemes are tolerant against authority compromise, and compromising of up to (N −2) authorities does not bring the whole system down. We provide detailed analysis on security and performance to show feasibility of the scheme Nameless Control and Nameless Control-F. We firstly implement the real toolkit of a multi authority based encryption scheme Nameless Control and Nameless Control-F.

4. System Design



From this, architecture it consists of the following, data owner, data consumer and

ISSN: 2454-9924

the cloud server. First the data owner access for public key to the authorities as well asthe data consumer also access private key to the authorities.

5. Methodology

5.1 System Model

In our system, there are four types of entities: N Attribute Authorities (denoted as A), Cloud Server, Data Owners and Data Consumers. A user can be a Data Owner and a Data Consumer simultaneously. Authorities are assumed to have powerful computation abilities, and they are supervised by government offices because some attributes partially contain users‘ personally identifiable information. The whole attribute set is divided into N disjoint sets and controlled by each authority, therefore each authority is aware of only part of attributes. A Data Owner is the entity who wishes to outsource encrypted data file to the Cloud Servers. The Cloud Server, who is assumed to have adequate storage capacity, does nothing but store them. Newly joined Data Consumers request private keys from all of the authorities, and they do not know which attributes are controlled by which authorities. When the Data Consumers request their private keys

ISSN: 2454-9924

from the authorities, authorities jointly create corresponding private key and send it to them. All Data Consumers are able to download any of the encrypted data files, but only those whose private keys satisfy the privilege tree Tp can execute the operation associated with privilege p. The server is delegated to execute an operation p if and only if the user‘s credentials are verified through the privilege tree Tp Design Goals

Our goal is to achieve a multi-authority CP-ABE which: achieves the security defined above; guarantees the confidentiality of Data Consumers‘ identity information; and tolerates compromise attacks on the authorities or the collusion attacks by the authorities. For the visual comfort, we frequently use the following notations hereafter. Ak denotes the k-th attribute authority; Au denotes the attributes set of user u; Auk denotes the subset of Au controlled by Ak; and ATp denotes the attributes set included in tree Tp.

5.2 Nameless control construction

Setup At the system initialization phase, any one of the authorities chooses a bilinear group G0 of prime order p with generator g and publishes it. Then, all authorities

ISSN: 2454-9924

independently and randomly picks vk ∈ Zp and send Yk = e(g, g)vk to all ther authorities who individually compute Y :=

_k∈A Yk = e(g, g)_k∈A vk . Then, every authority Ak randomly picks N − 1 integers skj ∈ Zp(j ∈ {1, . . . , N}\{k}) and computes gskj . Each gskj is shared with each other authority Aj. An authority Ak, after receiving N −1 pieces of gs jk generated by Aj.

We have assumed semi-honest authorities in Nameless Control and we assumed that they will not collude with each there. This is a necessary assumption in Nameless Control because each authority is in charge of a subset of the whole attributes set, and for the attributes that it is in charge of; it knows the exact information of the key requester. If the information from all authorities is gathered altogether, the complete attribute set of the key requester is recovered and thus his identity is disclosed to the authorities. In this sense, Nameless Control is semi anonymous since partial identity information (represented as some attributes) is disclosed to each authority, but we can achieve a full-anonymity and also allow the collusion of the authorities. The key point of the identity information leakage we had in our previous

ISSN: 2454-9924

scheme as well as every existing attribute based encryption schemes is that key generator (or attribute authorities in our scheme) issues attribute key based on the reported attribute, and the generator has to know the user‘s attribute to do so. We need to introduce a new technique to let key generators issue the correct attribute key without knowing what attributes the users have. A naive solution is to give all the attribute keys of all the attributes to the key requester and let him pick whatever he wants. In this way, the key generator does not know which attribute keys the key requester picked, but we have to fully trust the key requester that he will not pick any attribute key not allowed to him. To solve this, we leverage the following Oblivious Transfer (OT).

5.3 Fully Anonymous Multi-Authority CP-ABE

In this section, we present how to achieve the full anonymity in Nameless Control to designs the fully anonymous privilege control scheme Nameless Control -F. The Key Generate algorithm is the only part which leaks identity information to each attribute authority. Upon receiving the attribute key request with the attribute value, the attribute authority will generate H(att (i))ri and sends it to the requester where att (i) is the attribute value and ri is a random number for that attribute. The attribute value is disclosed to the authority in this step. We can introduce the above 1-out-of-n OT to prevent this leakage. We let each authority be in charge of all attributes belonging to the same category. For each attribute category c (e.g., University), suppose there are k possible attribute values (e.g., IIT, NYU, CMU ...), then one requester has at most one attribute value in one category. Upon the key request, the attribute authority can pick a random number ru for the requester and generates H(att (i)) ru for all i ∈ {1, . . . , k}. After the attribute keys are ready, the attribute authority and the key requester are engaged in a 1-out-of-k OT where the key requester wants to receive one attribute key among k. By introducing the 1-out-of-k OT in our Key Generate algorithm, the key requester achieves the correct attribute key that he wants, but the attribute authority does not have any useful information about what attribute is achieved by the requester. Then, the key requester achieves the full anonymity in our scheme and no matter how

ISSN: 2454-9924

many attribute authorities collude; his identity information is kept secret.

6. Conclusion

From this, N joint Authorities Integration of secured cloud data access privilege with attribute based encryption over distributed authorities has been implemented. This paper proposes a semi-anonymous attribute-based privilege control scheme Nameless Control and a fully-anonymous attribute-based privilege control scheme Nameless Control-F to address the user privacy problem in a cloud storage server. Using multiple authorities in the cloud computing system, our proposed schemes achieve not only fine-grained privilege control but also identity anonymity while conducting privilege control based on users‘ identity information. More importantly, our system can tolerate up to N − 2 authority compromise, which is highly preferable especially in Internet-based cloud computing environment. Additionally, we also enhance the system performance and efficiency of the system.

7. References

[1]A. Shamir, ―Identity-based cryptosystems and signature schemes,‖ in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 1985, pp. 47–53.

[2] A. Sahai and B. Waters, ―Fuzzy identity-based encryption,‖ in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 2005, pp. 457–473.

[3] V. Goyal, O. Pandey, A. Sahai, and B. Waters, ―Attribute-based encryption for fine-grained access control of encrypted data,‖ in Proc. 13th CCS, 2006, pp. 89–98.

[4] J. Bethencourt, A. Sahai, and B. Waters, ―Ciphertext-policy attributebased encryption,‖ in Proc. IEEE SP, May 2007, pp. 321–334.

[5] M. Chase, ―Multi-authority attribute based encryption,‖ in Theory of Cryptography. Berlin, Germany: Springer-Verlag, 2007, pp. 515–534.

[6] M. Chase and S. S. M. Chow, ―Improving privacy and security in multi-authority attribute-based encryption,‖ in Proc. 16th CCS, 2009, pp. 121–130.

[7] H. Lin, Z. Cao, X. Liang, and J. Shao,

ISSN: 2454-9924

-Secure threshold multi authority attribute

ISSN: 2454-9924

based encryption without a central authority,‖ Inf. Sci., vol. 180, no. 13, pp. 2618–2632, 2010.

[8] V. Božovi´c, D. Socek, R. Steinwandt, and V. I. Villányi, ―Multi-authority attribute-based encryption with honest-but-curious central authority,‖ Int. J. Comput. Math., vol. 89, no. 3, pp. 268–283, 2012.

[9] F. Li, Y. Rahulamathavan, M. Rajarajan, and R. C.-W. Phan, ―Low complexity multi-authority attribute based encryption scheme for mobile cloud computing,‖ in Proc. IEEE 7th SOSE, Mar. 2013, pp. 573–577.

[10] K. Yang, X. Jia, K. Ren, and B. Zhang, ―DAC-MACS: Effective data access control for multi-authority cloud storage systems,‖ in Proc. IEEE INFOCOM, Apr. 2013, pp. 2895–2903

ISSN: 2454-9924

Providing Security for Public Data Auditability in Regeneration based Cloud Storage

R. Saranya #1, S. Vanakovarayan *1

Mailam Engineering College, Mailam #1, *1

Saransweety08@gmail.com #1

Abstract: Storing data in a third party's cloud system provides serious anxiety across data privacy. To make sure remote data truth and fault tolerance often lacks the support of either public auditability or changing data operations. Task of allowing a trusted party auditor (TPA), on behalf of the cloud client, to verify the integrity of the dynamic data stored in the cloud. To provide efficient data dynamics, we improve the existing proof of storage models by implementing block tag authentication. Additionally, implement the technique of signature to extend our main result into a multi-user setting, where TPA can perform multiple auditing tasks simultaneously. And also we use a proxy agent to verify the signature technique and to determine the regeneration problem in case of failed authenticators and data owner.

Keywords: Third Party Auditor, Anxiety, Signature, Tolerance

Introduction

Cloud storage is now obtaining popularity because it offers a flexible on-demand data outsourcing service with providing benefits: relief of the load expenses for storage management, universal data access with location independence, and avoidance of capital expenditure on hardware, software, and personal maintenances, etc. However, this new paradigm of data hosting service also brings new security threats toward user's data, thus making individuals or enterprisers still feel hesitant. It is noted that data owners lose ultimate control over the fate of their outsourced data; thus, the correctness, availability and integrity of the data are being put at risk. On the one hand, the cloud service is usually faced with a

ISSN: 2454-9924

broad range of internal/external adversaries, who would maliciously delete or corrupt users' data; on the other hand, the cloud service providers may act dishonestly, attempting to hide data loss or corruption and claiming that the files are still correctly stored in the cloud for reputation or monetary reasons. Thus it makes great sense for users to implement an efficient protocol to perform periodical verifications of their outsourced data to ensure that the cloud indeed maintains their data correctly. Many mechanisms dealing with the integrity of outsourced data without a local copy have been proposed under different system and security models up to now. The most significant work among these studies is the PDP (provable data possession) model and

ISSN: 2454-9924

POR (proof of irretrievability) model, which were originally proposed for the single-server scenario. Considering that files are usually striped and redundantly stored across multi-servers or multi-clouds, explore integrity verification schemes suitable for such multi-servers or multi clouds setting with different redundancy schemes, such as replication, erasure codes, and, more recently, regenerating codes. In this paper, we focus on the integrity verification problem in regenerating-code-based cloud storage, especially with the functional repair strategy. Similar studies have been performed separately and independently. Extended the single-server CPOR scheme(private version in to the regenerating code-scenario; designed and implemented a data integrity protection(DIP) scheme for FMSR based cloud storage and the scheme is adapted to the thin-cloud setting. However, both of them are designed for private audit, only the data owner is allowed to verify the integrity and repair the faulty servers. Considering the large size of the outsourced data and the user's constrained resource capability, the tasks of auditing and reparation in the cloud can be formidable and expensive. The overhead of using cloud storage should be minimized as much as possible such that a user does not need to perform too many operations to their outsourced data (in additional to retrieving it). In particular, users may not want to go through the complexity in verifying and reparation. The auditing schemes in imply the problem that users need to always stay online, which may impede its adoption in practice, especially for long-term archival storage. To fully ensure the data integrity and save the users' computation resources as well as online burden, we propose a public auditing

ISSN: 2454-9924

scheme for the regenerating-code-based cloud storage, in which the integrity checking and regeneration (of failed data blocks and authenticators) are implemented by a third party auditor and a semi-trusted proxy separately on behalf of the data owner. Instead of directly adapting the existing public auditing scheme to the multi- server setting, we design a novel authenticator, which is more appropriate for regenerating codes. Besides, we "encrypt" the coefficients to protect data privacy against the auditor, which is more lightweight than applying the proof blind technique and data blind method. Several challenges and threats spontaneously arise in our new system model with a proxy, and security analysis shows that our scheme works well with these problems.

Related Work

Although the existing schemes aim at providing integrity verification for different data storage systems, the problem of supporting both public auditability and data dynamics has not been fully addressed. It does not support fault tolerance in case of failed authenticators. How to achieve a secure and efficient design to seamlessly integrate these two important components for data storage service remains an open challenging task in Cloud Computing.

Drawbacks

- Especially to support block insertion, which is missing in most existing schemes.

ISSN: 2454-9924

- Authentication was provided only at the time of upload not at the time of download.
- Data integrity is not maintained.
- Fault tolerance is not provided.
- Only single copy of data is stored.

Proposed Work

Here we are providing better security in owner's upload side as well as on the download side. For better security client splitting that single file into different blocks and providing a unique identification number for each block. Client: An entity, which has large data files to be stored in the cloud and relies on the cloud for data maintenance and computation, can be either individual consumers or organizations. Cloud Storage Server (CSS): An entity, which is managed by Cloud Service Provider (CSP), has significant storage space and computation resource to maintain the clients' data. Trusted Party Auditor (TPA): An entity, which has expertise and capabilities that clients do not have, is trusted to assess and expose risk of cloud storage services on behalf of the clients upon request. Proxy agent: semi-trusted and acts on behalf of the data owner to regenerate authenticators and data blocks on the failed servers during the repair procedure

Benefits

- We motivate the public auditing system of data storage security in Cloud Computing, and propose a protocol supporting for fully

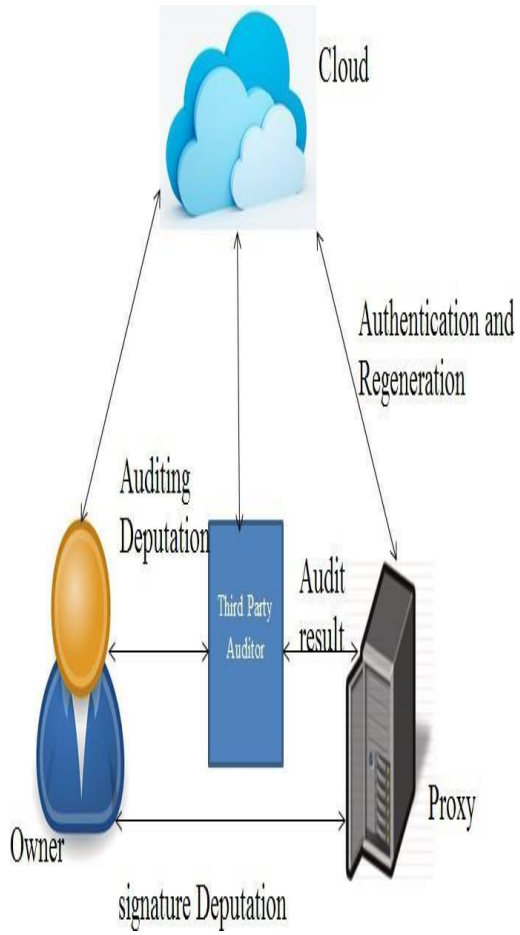
ISSN: 2454-9924

dynamic data operations, especially to support block insertion, which is missing in most existing schemes.

- We extend our scheme to support scalable and efficient public auditing in Cloud Computing. In particular, our scheme achieves auditing tasks from different users can be performed simultaneously by the TPA.
- Spitted into blocks and it is uploaded in the cloud in a encrypted format for better security.
- We prove the security of our proposed construction and justify the performance of our scheme through concrete implementation and comparisons.

System Design

ISSN: 2454-9924



Components involved in system design

- Data owner
- Cloud
- Third party Auditor(tpa)
- Proxy agent

The data owner owns large amounts of data files to be stored in the cloud. And the cloud which is managed by the cloud service provider, provide storage service and have significant computational resources. The

ISSN: 2454-9924

third party auditor who has expertise capabilities to conduct public audits on the coded data in the cloud, the TPA is trusted and its audit result is unbiased for both data owners and cloud servers. And a proxy agent who is semi trusted and acts on behalf of the data owner to regenerate authenticators and data blocks on the failed servers during the repair procedure. The data owner is restricted in computational and storage resources compared to other entities and may become offline even after the upload procedure. The TPA and proxy are much more powerful than the data owner but less than the cloud servers in terms of computation and memory capacity. To save resources as well as the online burden potentially brought by the periodic auditing and accidental repairing, the data owners resort to the TPA for integrity verification and delegate the reparation to the proxy.

Algorithm and description

The user can also encrypt data before outsourcing it into the cloud server with encryption techniques .As a significant research area for system protection, data access control has been evolving in the past thirty years and various techniques have

ISSN: 2454-9924

been developed to effectively implement fine-grained access control, which allows flexibility in specifying differential access rights of individual users. Traditional access control architectures usually assume the data owner and the servers storing the data are in the same trusted domain, where the servers are fully entrusted as an omniscient reference monitor responsible for defining and enforcing access control policies. This assumption however no longer holds in cloud computing since the data owner and cloud servers are very likely to be in two different domains. With the character of low maintenance, cloud computing provides an economical and efficient solution for sharing group resource among cloud users. Unfortunately, sharing data in a multi-owner manner while preserving data and identity privacy from an untrusted cloud is still a challenging issue, due to the frequent change of the membership. In this paper, we propose a secure multi-owner data sharing scheme, named Mona, for dynamic groups in the cloud. By leveraging group signature and dynamic broadcast encryption techniques, any cloud user can anonymously share data with others. Meanwhile, the storage overhead and encryption

ISSN: 2454-9924

computation cost of our scheme are independent with the number of revoked users. In addition, we analyze the security of our scheme with rigorous proofs, and demonstrate the efficiency of our scheme in experiments.

Ring signature:

Holomorphic Authenticators Ring Signatures Rijndael Managed, Suppose that a group of entities each have public/private key pairs, (PK1, SK1), (PK2, SK2),

...

,(PKn, SKn). Party i can compute a ring signature σ on a message m, on input (m, SKi, PK1, ... , PKn). Anyone can check the validity of a ring signature given σ, m, and the public keys involved, PK1, ... , PKn. If a ring signature is properly computed, it should pass the check. On the other hand, it should be hard for anyone to create a valid ring signature on any message for any group without knowing any of the secret keys for that group. Ring signatures, rst introduced by Rivest, Shamir, and Tauman, enable a user to sign a message so that a ring of possible signers (of which the user is a member) is identified, without revealing exactly which member of that ring actually generated the signature. In contrast to group

ISSN: 2454-9924

signatures, ring signatures are completely ad-hoc" and do not require any central authority or coordination among the various users (indeed, users do not even need to be aware of each other); furthermore, ring signature schemes grant users ne-grained control over the level of anonymity associated with any particular signature. This paper has two main areas of focus. First, we examine previous definitions of security for ring signature schemes and suggest that most of these prior definitions are too weak, in the sense that they do not take into account certain realistic attacks. We propose new definitions anonymity and un-forget ability which address these threats, and give separation results proving that our new notions are strictly stronger than previous ones. Second, we show the rst constructions of ring signature schemes in the standard model. One scheme is based on generic assumptions and satisfies our strongest definitions of security. Two additional schemes are more recent, but achieve weaker security guarantees and more limited functionality.

Methodology

- User registration
- Mail alert process
- Block insertion or block tag

1. User registration

Users can use them from anywhere at any time. For example, the email service is probably the most popular one. Cloud computing is a concept that treats the resources on the Internet as a unified entity, a cloud. Users just use services without being concerned about how computation is done and storage is managed. In this paper, we focus on designing a cloud storage system for robustness, confidentiality, and functionality. A cloud storage system is considered user interface entry level creation in this module. A user should register their details first such as their name, email id, mobile no, and an id for that particular client this is used to avoid the duplication of entries in this application.

2. Mail alert process

The uploading process of the user is first get the secret key in the corresponding user email id and then apply the secret key to

ISSN: 2454-9924

encrypted data to send the server storage and

ISSN: 2454-9924

decrypts it by using his secret key to download the corresponding data file in the server storage system's the secret key conversion using the Share Key Gen (SKA, t, m). This algorithm shares the secret key SKA of a user to a set of key servers, this key was generated and sent to the registered client's mail id as per the notifications which I have received from your end.

3. Block insertion or Block tag

Splitting the source content into 9 different blocks and have to give a tag for that all blocks for client identification, here that user defined tags are going to convert in to a binary value that's via ASCII table using Horner method. Dividing of these blocks makes difficult for attacker to predict the combinations also and the generated ASCII value acts as a metadata key value and each block is send along with the metadata key.

Conclusion

From this the providing preserving public auditing for regenerating code based cloud storage has been implemented. Where the data owners are privileged to delegate TPA

ISSN: 2454-9924

for checking entitled their data validity. To protect the original data privacy against the TPA, we randomize the coefficients in the beginning rather than applying the blind technique during the auditing process. Determining that the data owner cannot always stay online in practice, in order to keep the storage available and verifiable after a malicious corruption, we introduce a semi-trusted proxy into the system model and provide a privilege for the proxy to handle the reparation of the coded blocks and authenticators. To better exact for the regenerating-code-scenario, we design our authenticator based on the BLS signature. This authenticator can be efficiently generated by the data owner simultaneously with the encoding procedure. Additional analysis shows that our scheme is provable secure, and the performance evaluation shows that our scheme is highly effective t and can be feasibly integrated into a regenerating-code-based cloud storage system.

References

[1] A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, “Above the clouds: A

ISSN: 2454-9924

Berkeley view of cloud computing," Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS, vol. 28, p. 13, 2009.

[2] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrusted stores," in Proceedings of the 14th ACM Conference on Computer and Communications Security, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 598–609.[3] A. Juels and B. S. Kaliski Jr, "Pors: Proofs of retrievability for large files," in Proceedings of the 14th ACM conference on Computer and communications security. ACM, 2007, pp. 584–597.

[4] R. Curtmola, O. Khan, R. Burns, and G. Ateniese, "Mr-pdp: Multiplereplica provable data possession," in Distributed Computing Systems, 2008. ICDCS'08. The 28th International Conference on. IEEE, 2008, pp. 411–420.

[5] K. D. Bowers, A. Juels, and A. Oprea, "Hail: a high-availability and integrity layer for cloud storage," in Proceedings of the 16th ACMconference on Computer and communications security. ACM, 2009, pp. 187–198.

[6] J. He, Y. Zhang, G. Huang, Y. Shi, and J. Cao, "Distributed data possession checking for securing multiple replicas in geographically dispersed clouds," Journal of Computer and System Sciences, vol. 78, no. 5, pp. 1345–1358, 2012.

[7] B. Chen, R. Curtmola, G. Ateniese, and R. Burns, "Remote data checking for network coding-based distributed storage systems," in Proceedings of the 2010 ACM workshop on Cloud computing security workshop. ACM, 2010, pp. 31–42.

[8] H. Chen and P. Lee, "Enabling data integrity protection in regenerating coding-based cloud storage: Theory and implementation," Parallel and Distributed Systems, IEEE Transactions on, vol. 25, no. 2, pp. 407–416, Feb 2014.

ISSN: 2454-9924

RANDOMIZED DISPERSED STORAGE SYSTEM FOR MULTIPLE CLOUD WITH NETWORK CODING

G. Hemavathi,
M.E*, Dept. of CSE,
Dr. Pauls Engineering College,
ghemavathicse@gmail.com

Mrs. D. Udaya,
Assistant Professor,
Dept. of CSE,
Dr. Pauls Engineering College,

Abstract

In cloud storage data are striped across multiple cloud vendors to provide fault tolerance. When the cloud permanently fails due to any disaster, can be repaired using other surviving clouds to preserve data redundancy. In existing system they implemented a conventional erasure codes performs when some cloud experience short term transient failure not the permanent failure. In proposed, implements a proxy-based storage system for fault-tolerant multiple-cloud storage called NC Cloud, which achieves cost-effective repair for a permanent single-cloud failure. NC Cloud is built on top of a network-coding-based storage scheme called the functional minimum-storage regenerating (FMSR) codes, which maintain the same fault tolerance and data redundancy as in traditional erasure codes (e.g., RAID-6), but use less repair traffic and hence incur less monetary cost due to data transfer. This system implement a proof-of-concept prototype of NC Cloud and deploy it at top of both local and commercial clouds. By using FMSR, provides key feature to relax encoding requirement when the repair operation.

Keywords: Regenerating codes, network coding, fault tolerance, recovery,

1. INTRODUCTION

Cloud storage provides an on-demand remote backup solution. However, using a single cloud storage provider raises concerns such as having a single point of failure and vendor lock-ins. While striping data with conventional erasure

ISSN: 2454-9924

codes performs well when some clouds experience short-term transient failures or foreseeable permanent failures , there are real-life cases showing that permanent failures do occur and are not always foreseeable . In view of this, this work focuses on unexpected permanent cloud failures. When a cloud fails permanently, it is necessary to activate repair to maintain data redundancy and fault tolerance. A repair operation retrieves data from existing surviving clouds over the network and reconstructs the lost data in a new cloud. Today's cloud storage providers charge users for outbound data, so moving an enormous amount of data across clouds can introduce significant monetary costs. It is important to reduce the repair traffic (i.e., the amount of data being transferred over the network during repair), and hence the monetary cost due to data migration. To minimize repair traffic, regenerating codes have been proposed for storing data redundantly in a distributed storage system (a collection of interconnected storage nodes). Each node could refer to a simple storage device, a storage site, or a cloud storage provider. Regenerating codes are built on the concept of network coding, in the sense that nodes perform encoding operations and send encoded data. During repair, each surviving node encodes its stored data chunks and sends the encoded chunks to a new node, which then regenerates the lost data. It is shown that regenerating codes require less repair traffic than traditional erasure codes with the same fault tolerance level. Regenerating codes have been extensively studied in the theoretical context. However, the practical performance of regenerating codes remains uncertain. One key challenge for deploying regenerating codes in practice is that most existing regenerating codes require storage nodes to be equipped with computation capabilities for performing encoding operations during repair. On the other hand, to make regenerating codes portable to any cloud storage service, it is desirable to assume only a thin-cloud inter- face, where storage nodes only need to support the standard read/write functionalities.

2. IMPORTANCE OF REPAIR IN MULTIPLE- CLOUD STORAGE

We consider two types of failures: transient failure and permanent failure.

Transient failure: A transient failure is expected to be short-term, such that the "failed" cloud will return to normal after some time and no outsourced data is lost. We highlight that even though Amazon claims that its service is designed for providing 99.99% availability , there are arising concerns about this claim and the reliability of other cloud providers after Amazon's outage in April 2011 . We thus expect

ISSN: 2454-9924

that transient failures are common, but they will eventually be recovered. If we deploy multiple-cloud storage with enough redundancy, then we can retrieve data from the other surviving clouds during the failure period.

Permanent failure: A permanent failure is long-term, in the sense that the outsourced data on a failed cloud will become permanently unavailable. Clearly, a permanent failure is more disastrous than a transient one. Although we expect that a permanent failure is unlikely to happen, there are several situations where permanent cloud failures are still possible: • Data center outages in disasters. AFCOM [48] found that many data centers are ill-prepared for disasters. For example, 50% of the respondents have no plans to repair damages after a disaster. It was reported [48] that the earthquake and tsunami in northeastern Japan in March 11, 2011 knocked out several data centers there. • Data loss and corruption. There are real-life cases where a cloud may accidentally lose data [12], [40], [58]. In the case of Magnolia [40], half a terabyte of data, including its backups, are all lost and unrecoverable. • Malicious attacks. To provide security guarantees for outsourced data, one solution is to have the client application encrypt the data before putting the data on the cloud. On the other hand, if the outsourced data is corrupted (e.g., by virus or malware), then even though the content of the data is encrypted and remains confidential to outsiders, the data itself is no longer useful. AFCOM found that about 65 percent of data centers have no plan or procedure to deal with cyber-criminals.

3. MOTIVATION OF FMSR CODES

We consider a distributed, multiple-cloud storage set- ting from a client's perspective, where data is striped over multiple cloud providers. We propose a proxy- based design that interconnects multiple cloud repositories. The proxy serves as an interface between client applications and the clouds. If a cloud experiences a permanent failure.

We consider fault-tolerant storage based on a type of maximum distance separable (MDS) codes. Given a file object of size M, we divide it into equal-size native chunks, which are linearly combined to form code chunks. When an (n,k)-MDS code is used, the native/code chunks are then distributed over n (larger than k) nodes, each storing chunks of a total size M/k, such that the original file object may be reconstructed from the chunks contained in any k of the n nodes. Thus, it tolerates the failures of any n − k nodes. We call this fault tolerance feature the MDS property. The extra feature of FMSR codes is that reconstructing the chunks stored in a failed node can be achieved

ISSN: 2454-9924

by downloading less data from the surviving nodes than reconstructing the whole file. This paper considers a multiple-cloud setting with two levels of reliability: fault tolerance and recovery. First, we assume that the multiple-cloud storage is double- fault tolerant (e.g., as in conventional RAID-6 codes) and provides data availability under the transient unavailability of at most two clouds. That is, we set $k = n - 2$. Thus, clients can always access their data as long as no more than two clouds experience transient failures (see examples in Table 1) or any possible connectivity problems. We expect that such a fault tolerance level suffices in practice. Second, we consider single-fault recovery in multiple-cloud storage, given that a permanent cloud failure is less frequent but possible. Our primary objective is to minimize the cost of storage repair for a permanent single-cloud failure. In this work, we focus on comparing two codes: traditional RAID-6 codes and our FMSR codes with double-fault tolerance3. We define the repair traffic as the amount of outbound data being downloaded from the other surviving clouds during the single-cloud failure recovery. We seek to minimize the repair traffic for cost-effective repair. Here, we do not consider the inbound traffic (i.e., the data being written to a cloud), as it is free of charge for many cloud providers (see Table 3 in Section 6). We now study the repair traffic involved in different coding schemes via examples. Suppose that we store a file of size M on four clouds, each viewed as a logical storage node. Let us first consider conventional RAID-6 codes, which are double-fault tolerant. Here, we consider a RAID-6 code implementation based on the Reed-Solomon code. We divide the file into two native chunks (i.e., A and B) of size M/2 each. We add two code chunks formed by the linear combinations of the native chunks. Suppose now that Node 1 is down. Then the proxy must download the same number of chunks as the original file from two other nodes (e.g., B and A + B from Nodes 2 and 3, respectively). It then reconstructs and stores the lost chunk A on the new node. The total storage size is 2M, while the repair traffic is M. Regenerating codes have been proposed to reduce the repair traffic. One class of regenerating codes is called the exact minimum-storage regenerating (EMSR) codes . EMSR codes keep the same storage size as in RAID- 6 codes, while having the storage nodes send encoded chunks to the proxy so as to reduce the repair traffic. Figure 2(b) illustrates the double-fault tolerant implementation of EMSR codes. We divide a file into four chunks, and allocate the native and code chunks as shown in the figure. Suppose Node 1 is down. To repair it, each surviving node sends the XOR summation of the data chunks to the proxy, which then reconstructs the lost chunks. We can see that in EMSR codes, the storage size is 2M (same as RAID-6 codes), while the repair traffic is 0.75M,

ISSN: 2454-9924

which is 25% of saving (compared with RAID-6 codes). EMSR codes leverage the notion of network coding, as the nodes generate encoded chunks during repair.

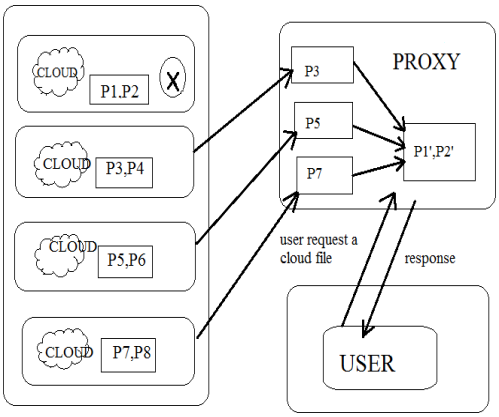


Figure: System model for Repair operation

4. FMSR CODE IMPLEMENTATION

We now present the details for implementing FMSR codes in multiple-cloud storage. We specify three operations for FMSR codes on a particular file object: (1) file upload; (2) file download; (3) repair. Each cloud repository is viewed as a logical storage node. Our implementation assumes a thin-cloud interface [60], such that the storage nodes (i.e., cloud repositories) only need to support basic read/write operations. Thus, we expect that our FMSR code implementation is compatible with today's cloud storage services. One property of FMSR codes is that we do not require lost chunks to be exactly reconstructed, but instead in each repair, we regenerate code chunks that are not necessarily identical to those originally stored in the failed node, as long as the MDS property holds. We propose a two-phase checking scheme, which ensures that the code chunks on all nodes always satisfy the MDS property, and hence data availability, even after iterative repairs. In this section, we analyze the importance of the two-phase checking scheme.

4.1 Basic operations

4.1.1 File Upload

To upload a file F, we first divide it into $k(n-k)$ equal- size native chunks, denoted by $(Fi)i=1,2,\cdots,k(n-k)$. We then encode these $k(n-k)$ native chunks into $n(n-k)$ code chunks, denoted by $(Pi)i=1,2,\cdots,n(n-k)$. Each Pi is formed by a linear combination of the $k(n-k)$ native chunks. Specifically, we let $EM = [\alpha i,j]$ be an $n(n-k)\times k(n-k)$ encoding matrix for some coefficients $\alpha i,j$ (where $i = 1,...,n(n-k)$ and $j = 1,...,k(n-k)$) in the Galois field GF(28). We call a row vector of EM an encoding coefficient

ISSN: 2454-9924

vector (ECV), which contains k(n−k) elements. We let ECVi denote the ith row vector of EM. We com- pute each Pi by the product of ECVi and all the native chunks F1,F2,⋯, Fk(n−k), i.e., Pi = Pk(n−k) j=1 αi,jFj for i = 1,2,⋯,n(n−k), where all arithmetic operations are performed over GF(28). The code chunks are then evenly stored in the n storage nodes, each having (n−k) chunks. Also, we store the whole EM in a metadata object that is then replicated to all storage nodes There are many ways of constructing EM, as long as it passes our two-phase checking. Note that the implementation details of the arithmetic operations in Galois Fields are extensively discussed in

4.1.2 File Download

To download a file, we first download the corresponding metadata object that contains the ECVs. Then we select any k of the n storage nodes, and download the k(n−k) code chunks from the k nodes. The ECVs of the k(n−k) code chunks can form a k(n−k)×k(n−k) square matrix. If the MDS property is maintained, then by definition, the inverse of the square matrix must exist. Thus, we multiply the inverse of the square matrix with the code chunks and obtain the original k(n − k) native chunks. The idea is that we treat FMSR codes as standard Reed-Solomon codes, and our technique of creating an inverse matrix to decode the original data has been described in the tutorial [46].

4.1.3 Iterative Repairs

We now consider the repair of FMSR codes for a file F for a permanent single-node failure. Given that FMSR codes regenerates different chunks in each repair, one challenge is to ensure that the MDS property still holds even after iterative repairs. This is in contrast to regenerating the exact lost chunks as in RAID-6, which guarantees the invariance of the stored chunks. Here, we propose a two-phase checking heuristic as follows. Suppose that the (r −1)th repair is successful, and we now consider how to operate the rth repair for a single permanent node failure (where r ≥ 1). We first check if the new set of chunks in all storage nodes satisfies the MDS property after the rth repair. In addition, we also check if another new set of chunks in all storage nodes still satisfies the MDS property after the (r + 1)th repair, should another single permanent node failure

ISSN: 2454-9924

occur (we call this the repair MDS (rMDS) property).

5. NCCLOUD DESIGN AND IMPLEMENTATION

We implement NC Cloud as a proxy that bridges user applications and multiple clouds. Its design is built on three layers. The file system layer presents NC Cloud as a mounted drive, which can thus be easily interfaced with general user applications. The coding layer deals with the encoding and decoding functions. The storage layer deals with read/write requests with different clouds. Each file is associated with a metadata object, which is replicated at each repository. The metadata object holds the file details and the coding information (e.g., encoding coefficients for FMSR codes). NCCloud is mainly implemented in Python, while the coding schemes are implemented in C for better efficiency. The file system layer is built on FUSE . The coding layer implements both RAID-6 and FMSR codes. Our RAID-6 code implementation is based on the Reed-Solomon code (as shown in Figure 2(a)) for baseline evaluation. We use zfec to implement the RAID-6 codes, and we utilize the optimizations made in zfec to implement FMSR codes for fair comparison. Recall that FMSR codes generate multiple chunks to be stored on the same repository. To save the request cost overhead, multiple chunks destined for the same repository are aggregated before upload. Thus, FMSR codes keep only one (aggregated) chunk per file object on each cloud, as in RAID-6 codes. To retrieve a specific chunk, we calculate its offset within the combined chunk and issue a range GET request.

6. EVALUATION

6.1 Cost Analysis

6.1.1 Repair Cost Saving

We first analyze the saving of monetary costs in repair in practice. Table 3 shows the monthly price plans for three major providers as of May 2013. We take the cost from the first chargeable usage tier (i.e., storage usage within 1TB/month; data transferred out more than 1GB/month but less than 10TB/month). From the analysis in Section 3, we can save 25-50% of the download traffic during storage repair.

ISSN: 2454-9924

The storage size and the number of chunks being generated per file object are the same in both RAID-6 and FMSR codes (assuming that we aggregate chunks in FMSR codes as described in). However, in the analysis, we have ignored two practical considerations: the size of metadata (Section 5) and the number of requests issued during repair. We now argue that they are negligible and that the simplified calculations based only on file size suffice for real-life applications. Metadata size: Our implementation currently keeps the metadata size of FMSR codes within 160 bytes when n = 4 and k = 2, regardless of the file size. For a large n, say when n = 12 and k = 10, the metadata size

10 is still within 900 bytes. NCCloud aims at long-term backups, and can be integrated with other backup applications. Existing backup applications typically aggregate small files into a larger data chunk in order to save the processing overhead. For example, the default setting for Cumulus creates chunks of around 4MB each. Thus, the metadata size overhead can be made negligible. Since both RAID-6 and FMSR codes store the same amount of file data, they incur very similar storage costs in normal usage (assuming that the metadata costs are negligible). Number of requests providers charge for requests. RAID-6 and FMSR codes differ in the number of requests when retrieving data during repair. Suppose that we store a file object of size 4MB with n = 4 and k = 2. During repair, RAID-6 and FMSR codes retrieve two and three chunks, respectively. The cost overhead due to the GET requests for RAID-6 codes is at most 0.171%, and that for FMSR codes is at most 0.341%, a mere 0.17% increase.

6.2 Response Time Analysis

We deploy our NCCloud prototype in real environments. We evaluate the response time performance of three basic operations, namely file upload, file download, and repair, in two scenarios. The first part analyzes in detail the time taken by different NCCloud operations. It is done on a local cloud storage tested in order to lessen the effects of network fluctuations. The second part evaluates how NCCloud actually performs in a commercial cloud environment. All results are averaged over 40 runs. We assume that repair coefficients are generated offline , so the time taken by

ISSN: 2454-9924

two-phase checking is not accounted for in the repair operation.

7. RELATED WORK

We review the related work in multiple-cloud storage and failure recovery. Multiple-cloud storage. There are several systems proposed for multiple-cloud storage. HAIL provides integrity and availability guarantees for stored data. RACS uses erasure coding to mitigate vendor lock- ins when switching cloud vendors. It retrieves data from the cloud that is about to fail and moves the data to the new cloud. Unlike RACS, NCCloud excludes the failed cloud in repair. Vukoli´c advocates using multiple independent clouds to provide Byzantine fault tolerance. DEPSKY [10] addresses Byzantine fault tolerance by combining encryption and erasure coding for stored data. Single-node failure recovery schemes that minimize the amount of data read (or I/Os) for XOR-based erasure codes. For example, authors of [62], [63] propose opti- mal recovery for specific RAID-6 codes and reduce the amount of data read by up to around 25% (compared to conventional repair that downloads the amount of orig- inal data) for any number of nodes. Note that our FMSR codes can achieve 25% saving when the number of nodes is four, and up to 50% saving if the number of nodes increases. Authors of propose an enumeration-based approach to search for an optimal recovery solution for arbitrary XOR-based erasure codes. Efficient recovery is recently addressed in commercial cloud storage systems. For example, new constructions of non-MDS era- sure codes designed for efficient recovery are proposed for Azure and Facebook . The codes used in trade storage overhead for performance, and are mainly designed for data-intensive computing. Our work targets the cloud backup applications. Minimizing repair traffic. Regenerating codes stem from the concept of network codin2g and minimize the repair traffic among storage nodes. They exploit the optimal trade-off between storage cost and repair traffic, and there are two optimal points.

8 .CONCLUSIONS

We present NCCloud, a proxy-based, multiple-cloud storage system that practically addresses the reliability of today’s cloud backup storage. NCCloud not only provides fault tolerance in storage, but also allows cost-effective

ISSN: 2454-9924

repair when a cloud permanently fails. NCCloud implements a practical version of the functional minimum storage regenerating (FMSR) codes, which regenerates new parity chunks during repair subject to the required degree of data redundancy. Our FMSR code implementation eliminates the en- coding requirement of storage nodes (or cloud) during repair, while ensuring that the new set of stored chunks after each round of repair preserves the required fault tolerance. Our NCCloud prototype shows the effectiveness of FMSR codes in the cloud backup usage, in terms of monetary costs and response times.

9. REFERENCES

[1] H. Abu-Libdeh, L. Princehouse, and **H.** Weatherspoon. RACS: A Case for Cloud Storage Diversity. In Proc. of ACM SoCC, 2010.

[2] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network Information Flow. IEEE Trans. on Information Theory, 46(4):1204– 1216, Jul 2000.

[3] Amazon. AWS Case Study: Backupify. http://aws.amazon.com/ solutions/case-studies/backupify/.

[4] Amazon. Case Studies. https://aws.amazon.com/solutions/case-studies/#backup. [5] Amazon Glacier. http://aws.amazon.com/glacier/.

[6] Amazon S3. http://aws.amazon.com/s3.

[7] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Kon- winski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. A View of Cloud Computing. Communications of the ACM, 53(4):50–58, 2010.

[8] Asigra. Case Studies. http://www.asigra.com/product/case-studies/. [9] AWS Service Health Dashboard. Amazon s3 availability event: July 20, 2008. http://status.aws.amazon.com/s3-20080720.html.

[10] A. Bessani, M. Correia, B. Quaresma, F. Andr´e, and P. Sousa. DEPSKY: Dependable and Secure Storage in a Cloud-of-Clouds. In Proc. of ACM EuroSys, 2011.

[11] K. D. Bowers, A. Juels, and A. Oprea. HAIL: A High-Availability and Integrity Layer for Cloud Storage. In Proc. of ACM CCS, 2009. [12] Business Insider. Amazon's Cloud Crash Disaster Permanently Destroyed Many Customers' Data. http://www.businessinsider.

ISSN: 2454-9924

com/amazon-lost-data-2011-4/, Apr 2011.

[13] B. Calder et al. Windows Azure Storage: A Highly Available Cloud Storage Service with Strong Consistency. In Proc. of ACM SOSP, 2011.

[14] B. Chen, R. Curtmola, G. Ateniese, and R. Burns. Remote Data Checking for Network Coding-Based code.

ISSN: 2454-9924

Real Time Fraud Detection System using Data mining Techniques

K. Nithya #1, E. Lavanya *1

Mailam Engineering College, Mailam #1, *1

Nithyaknov14@gmail.com #1, lavanyakarthikeyanvpm@gmail.com *1

Abstract - Intelligent fraud detection system using random forest algorithm, detects the fraudulent card during transactions and alerts the customer regarding the fraud. This project also aims in minimizing the number of false alerts. The concept of random forest algorithm is a novel one in this application domain. The algorithm Improvements in classification accuracy have resulted from growing an ensemble of trees and letting them vote for the most popular class. The random forest algorithm used to detect the fraud detection in credit system and also enhance the system efficiency.

I. Introduction

Fraud refereed as to gaining goods/services and money by banned way. Fraud determines with events which involve criminal motives that, mostly, are hard to determine. Credit cards are one of the most popular aim of fraud but not the only one. Credit card fraud, a high range term for theft and fraud devoted or any similar payment mechanism as a fraudulent resource of funds in a transaction. Credit card fraud has been increasing issue in the credit card industry. Finding credit card fraud is a difficult task when using normal process, so the development of the credit card fraud detection models has become of importance whether in the academic or business organizations currently. In recent years, the prevailing data mining concerns people with credit card fraud detection model based on data mining. Since our problem is approached as a classification problem, classical data mining algorithms are not directly applicable. Intelligent fraud detection system using random forest algorithm, detects the fraudulent card during transactions and alerts the customer regarding the fraud. This project also aims in minimizing the number of false alerts. The concept of random forest algorithm is a novel one in this application domain. The algorithm Improvements in classification accuracy have resulted from growing an ensemble of trees and letting them vote for the most popular class. The random forest algorithm used to detect the fraud detection in credit system and also enhance the

system efficiency. It is documented in such way that, it is convenient to the user. Each section is divided into sub-sections. This chapter gives the information regarding analysis done for the proposed system. Here the goal of the project is explained, and also the cost and performance factors which will affect the feasibility of the project is explained, gets through the functional and nonfunctional requirement phase of the proposed system. This chapter illustrates the overall structure and responsibility of the project using UML. This chapter gets through the requirement phase of the proposed system and studies the requirements of the system in detail. It presents a formal document that crystallizes the user's requirements. The result of this study is being used in all the future steps of development of the project. In this chapter the detailed system design explores architecture of the system. It deals with the modules and their relationship in building the whole system. Design at this level explains about sub systems which are building blocks of the whole system. These sub systems have their well-defined functionality. The coding logic of the tools is explained with the code and syntax. We present how the code is organized with comments on code for understanding in future reference. We also discuss the Naming conventions that were followed during the Implementation phase of the project and also the descriptions of the methods of all the modules used by the system. This chapter gives the conclusion of the report and also the possible enhancements that could be done in the future.

II. Related Work

The Traditional detection method mainly depends on database system and the education of customers, which usually are delayed, inaccurate and not in-time. After that methods based on decimate analysis and regression analysis are widely used which can detect fraud by credit rate for cardholders and credit card transaction. For a large amount of data it is not efficient.

2.1 Proposed Method

The proposed system overcomes the above mentioned issue in an efficient way. This proposed system, aims in minimizing the number of false alerts. The concept of random forest algorithm is a novel one in this application domain. The algorithm Improvements in classification accuracy have resulted from growing an ensemble of trees and letting them vote for the most popular class. The random forest algorithm used to detect the fraud detection in credit system and also enhance the system efficiency

III. Literature Review

[1] Describes about, Credit card business has developed quickly in the world in the past years. In order to prevent defaulting risk, some useful tools are used in credit card management. Application Scoring and Behavior Scoring are two important steps in this process, which data mining algorithms are used widely. In this paper, the application of data mining in credit card management are analyzed and the experiments show the difference of data mining in the application scoring and behavior scoring. The amount of issued credit cards has increased rapidly in Taiwan and is characterized as high risk business if comparing to that of the traditional banking loan. To minimize the operating risks and maximize business profits, the credit card issuing institutions need an intelligent system to support the process of the risk management after cards issued. The aim of this study is to construct an efficient risk prediction system to detect the possible defaults for the credit card holders. The system collects the personal and financial information about the credit card holders and then applies evolutional neural network which integrated with grey incidence analysis and Dempster-Shafer theory of evidence to predict the default cases. The experimental results show the integrated model has better prediction accuracy if compare to the model which applies evolutional neural network only and is capable of tracing and reducing the default risks.

[2] Describes about, Credit card fraud on the Internet is a serious and growing issue. Many criminals have hacked into merchant databases to obtain cardholder details enabling them to conduct fake transactions or to sell the details in the digital underground economy. The card brands have set up a standard called PCI DSS to secure credit card details when they are stored online. We investigate the standard and find significant flaws especially in its requirements on small businesses. Finally, we propose some general rules for the secure management of online data. [18] Describes about In this paper, we study how to switch the customers from an undesirable class to a desirable one in credit card churning management by post mining. Multiple Criteria Linear Programming (MCLP) classification model, an optimization-based data mining method, is firstly used to classify the samples. In post mining phase, we build a case base formed by a series of typical positive instances for the entire negative population as their "good examples". These positive instances are on or near the boundary between the two classes, and thus closest to negative objects to ensure lowest switching cost. Switching

plan for each negative object is then generated based on the case base, according to minimum cost principle. Real dataset from a large commercial bank of China is used to validate the method we proposed.

[16] Describes about First, we classify the selected customers into clusters using RFM model to identify high-profit, gold customers. Subsequently, we carry out data mining using association rules algorithm. We measure the similarity, difference and modified difference of mined association rules based on three rules, i.e. emerging pattern rule, unexpected change rule, and added/perished rule. In the meantime, we use rule matching threshold to derive all types of rules and explore the rules with significant change based on the degree of change measured. In this paper, we employ data mining tools and effectively discover the current spending pattern of customers and trends of behavioral change, which allow management to detect in a large database potential changes of customer preference, and provide as early as possible products and services desired by the customers to expand the clientele base and prevent customer attrition.

IV. Random forest algorithm

We present pseudo-code for the basic algorithm only, without the bounded fringe technique. The addition of a bounded fringe is straightforward, but complicates the presentation significantly. Candidate split dimension A dimension along which a split may be made. Candidate split point One of the first m structure points to arrive in a leaf. Candidate split A combination of a candidate split dimension and a position along that dimension to split. These are formed by projecting each candidate split point into each candidate split dimension. Candidate children Each candidate split in a leaf induces two candidate children for that leaf. These are also referred to as the left and right child of that split.

V. Modules

5.1 Registration of user details

In this module, user have to register their detail first, the details are user id, name, date of birth, account number, balance and transaction details, and address.

5.2 Login of user

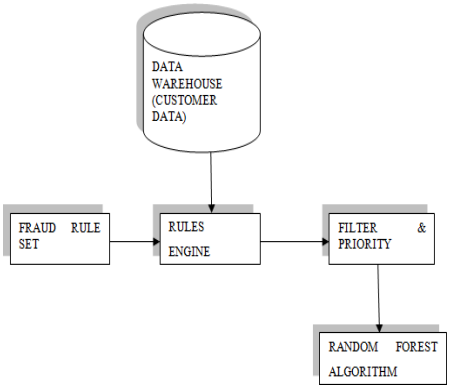
In this module, user have login with their own id password and pin number, then only he or she will access the process. The user details are maintained in the database for future references.

5.3 Brute force attack for password guessing

In this module describes about, it will show alert when the unauthorized person access the account or something wrong about account.

VI. Architecture

The above architecture describes the work structure of the system.The customer data in the data warehouse is subjected to the rules engine which consists of the fraud rule set. The filter and priority module sets the priority for the data and then sends it to the genetic algorithm which performs its functions and generates the output.



VII. Conclusion

From this, intelligent fraud detection system using random forest algorithm has been implemented. This project also aims in minimizing the number of false alerts. The concept of random forest algorithm is a novel one in this application domain. In future, the system will achieve more efficiency by improving the performances of the algorithm.

References

[1] Aihua Li , 1.Study on the Application of Data Mining Algorithms in Credit Card Management, Author(s) Aihua Li, 2009

[2] Blackwell.C,The management of online credit card data using the Payment Card Industry Data Security Standard, 2008

[3] Credit card fraud detection using hidden markov model – Abinav Srivastava,Amlan Kundu,Shamik Sural,Arun K.majumdar

[4] clifton phua, vincent lee1, kate smith & ross gayler, A Comprehensive Survey of Data Mining-based Fraud Detection Research,2005.

[5] Elio Lozano, Edgar Acu˜na, Parallel algorithms for distance-based and density-based outliers,2006.

[6] Leila Seyedhossein, Mahmoud Reza Hashemi Mining Information from Credit Card Time Series for Timelier Fraud Detection International Symposium on Telecommunications 2010.

[7] Linda Delamaire (UK), Hussein Abdou (UK), John Pointon (UK), “Credit card fraud and detection techniques: a review”2009.

[8] Md Delwar Hussain Mahdi, Karim Mohammed Rezaul, Muhammad Azizur Rahman “Credit Fraud Detection in the Banking Sector in UK: A Focus on E-Business.” 2010.

[9] M. Hamdi Ozcelik, Ekrem Duman, Mine Isik, Tugba Cevik, Improving a credit

card fraud detection system using genetic algorithm, International conference on Networking and information technology 2010.

[10] Mirjana Pejic-Bach, “Profiling intelligent systems applications in fraud detection and prevention: survey of research articles”, 2010.

[11] M.F. Gadi, X. Wang, and A.P. Lago, "Comparison with parametric optimization in credit card fraud detection, 2008.

[12] Raghavendra Patidar, Lokesh Sharma, “Credit Card Fraud Detection Using Neural Network” 2011.

[13] S. Benson Edwin Raj, A. Annie Portia “Analysis on Credit Card Fraud Detection Methods” 2011.

[14] Sahin, Y., Duman, E.: An overview of business domains where fraud can take place, and a survey of various fraud detection techniques. In: Proceedings of the 1st International Symposium on Computing in Science and Engineering, Aydin, Turkey (2010).

[15] Tej Paul Bhatla, Vikram Prabhu & Amit Dua “Understanding Credit Card Frauds,” 2003

[16] Ruey-Chyi Wu, Data mining application in customer relationship management of credit card business, 2005

[17] Wen-Fang YU, Na Wang, Research on Credit Card Fraud Detection Model Based on Distance Sum, IEEE International Joint Conference on Artificial Intelligence 2009.

[18] Yibing Chen , Post Mining of Multiple Criteria Linear Programming Classification Model for Actionable Knowledge in Credit Card Churning Management, 2011

ISSN: 2454-9924

Role Check Secured User Data Access with Attribute-based Encryption, Dynamic Key Generation & User Revocation System

D. Arthy #1, T. Priya Rathika Devi *1

Mailam Engineering College, Mailam #1, *1

dhakshina.arthy@gmail.com #1

Abstract

Attribute-based encryption (ABE) describes a mechanism for complex process control over secured data. The previous method only store the data in the cloud any user can access the data. In the proposed system the data is Stored in the Remote Cloud. Data Owner can share the Data and it's Key to the Permitted Users. Data Sharing is achieved for three types of Users. 1. User Based 2. Role Based (Position / Role), 3. Attribute (Experience). In the modified concept of the Project is Data is uploaded by the Data Owner based on Public Key, Secret Key, Global Key and Group Key. Public Key is randomly generated. Secret Key & Group Key is generated via our Role / Attribute. For both User Name & Designation is used. Global Key is generated randomly. We encrypt the uploaded file using AES Algorithm. User Revocation is also developed in this Part. If a user is moving out of the Group or removed out of the Group, Key is altered and the new Key is mailed to the Present Members. During Data Download, apart from verifying all the Keys, Token is generated as E Mail, which is used for further Authentication.

Key Words: Group key, Attribute, Encryption, Randomly, Secret key

1 Introduction

CLOUD computing is a very fascinating computing paradigm, in which computation and storage are moved away from terminal devices to the cloud. This new and popular paradigm brings important revolutions and makes bold innovations for the manner in which enterprises and individuals manage, distribute, and share content. By outsourcing their information technology capabilities to some cloud service providers, cloud users may achieve significant cost savings. There is widespread public concern about cloud computing, that is, how to ensure cloud users' data security. Part of the outsourced data is sensitive, should be accessed by

ISSN: 2454-9924

authorized data consumers at remote locations. For instance, in a university, one of its colleges uploads its (encrypted) development projects to the university cloud, and wants to give only the administrative personnel of the university and the faculty of this college the privilege to access the encrypted projects. This is a very natural scenario in our real life when we use cloud storage systems. Cryptographic technology is an essential manner to achieve this goal. For example, Attribute-Based Encryption (ABE), a special kind of powerful functional encryptions, contributes to access control over encrypted data. The notion of ABE was first introduced by Sahai and Waters. Depending on how to deploy the access control policy, there are two different kinds of ABE systems. That is, Key-Policy Attribute-Based Encryption (KP-ABE) and its dual notion, Cipher text-Policy Attribute-Based Encryption (CP-ABE). In a CP-ABE scheme, every cipher text is associated with an access policy, and every user's private key is associated with a set of attributes. While in a KPABE scheme, cipher texts are labeled with sets of attributes and access policies over these attributes are associated with users' private keys. In an ABE system, decryption operation requires that the set of attributes should match the access policy. Given its expressiveness, ABE is regarded as one of the most natural and important technologies for realizing data access control in the cloud. However, in most existing ABE schemes, one of the main efficiency drawbacks is that the size of the cipher text and the decryption overhead (computational cost) grow with the complexity of the access policy. This becomes critical barriers in applications running on resource-limited devices. For instance, the college adopts a pairing-based ABE scheme to encrypt its development projects and uploads the generated ABE cipher text to the university cloud. An authorized administrative officer of the university, who is on a business ship, wants to look up the encrypted development projects of the college through his (resource-limited) mobile phone. He then wants to download and decrypt the ABE cipher text. Since the cipher text might have a large size and the pairing operations in decryption procedure are usually expensive for a resource-limited device, he has to wait for a long time and sometimes even aborts the decryption procedure. To remarkably eliminate the cipher text size and the

ISSN: 2454-9924

decryption overhead for users in ABE systems, a new method for efficiently and securely outsourcing decryption of ABE cipher texts was put forth by Green et al. Fig. 1 illustrates their ABE system with outsourced decryption. More concretely, in their ABE system with outsourced decryption, the key generation algorithm is modified to produce two keys for a user. The first one is a short El Gama type secret key, called the retrieving key rk, and it must be kept private by the user.

2 Related Work

The existing concept only store the data in the cloud any user can access the data. Here it provides some drawback to the existing concept. They are, Congestion occurring, Less security, Less effective, Low connectivity

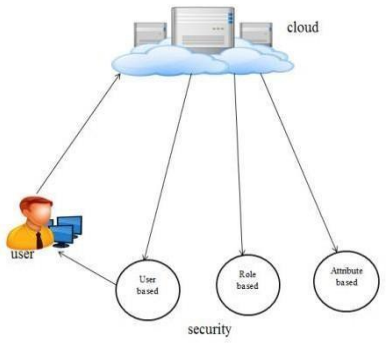
3 Proposed Work

In the proposed work the data is Stored in the Remote Cloud. Data Owner can share the Data and it's Key to the Permitted Users. Data Sharing is achieved for three types of Users. 1. User Based 2. Role Based (Position / Role), 3. Attribute (Experience). In the modified work of the Project is Data is uploaded by the Data Owner based on Public Key, Secret Key, Global Key and Group Key. Public Key is randomly generated. Secret Key & Group Key is generated via our Role / Attribute. For both User Name & Designation is used. Global Key is generated randomly. We encrypt the uploaded file using AES Algorithm. User Revocation is also developed in this Part. If a user is moving out of the Group or removed out of the Group, Key is altered and the new Key is mailed to the Present Members. During Data Download, apart from verifying all the Keys, Token is generated as E Mail, which is used for further Authentication. I this, it provides some benefits are,

- Avoid Congestion
- User friendly
- High security
- More effective

4 Architecture

ISSN: 2454-9924



5 Methodologies

5.1 User Registration

If the user wants to access the data from the server, they should have an account with that server. Without having an account they aren't able to access the files are view the details. So first the user will create an account with that server by providing the necessary information like Username, Password, DOB, Address and Phone number. Once this information is provided by the user, server will get that information and stored it into the database for future purpose.

5.2 Cloud Server

Cloud Data Service Provider will contain the large amount of data in their Data Storage. Also the Cloud Service provider will maintain the all the User information to authenticate the User when are login into their account. The User information will be stored in the Database of the Cloud Service Provider. Also the Data Server will redirect the User requested job to the Resource Assigning Module to process the User requested Job. The Request of all the Users will process by the Resource Assigning Module. To communicate with the Client and the with the other modules of the Network, the Data Server will establish connection between them. For this Purpose we are going to create a User Interface Frame. Also the Cloud Service Provider will send the User Job request to the Resource Assign Module in Fist in First out (FIFO) manner.

5.3 Data upload with Data sharing

Although the Cloud Computing is vast developing technology, In security point of view the it need more growth. To overcome this disadvantage, we implementing two types of Cloud. Once is Public Cloud and another one is Private Cloud. In Private the patient will set the access privileges' for each and every user they wish. In Public Cloud, the Cloud Server will set the access privileges' for each and every user based on their designation. So that legitimate users

ISSN: 2454-9924

can view the data stored in the cloud only up to their privilege level. They aren't allowed to view the data beyond their privileges'.

5.4 Three Layer User Access Control

In the three layer user access control system, the data is stored in the remote cloud. Data owner can share the data and its key to the permitted users. Data sharing is achieved for three types of users 1.User Based, 2.Role Based (Position/Role), 3.Attribute (Experience)

5.5 Request with Two Third Authentications

In this module is to share the data across the user using multiparty two third authentication schemes. Using this scheme new user can send the request to the data owner as well as permitted users. Either owner or two third o permitted user authenticates (SMS alert to the owner) the request, data is forwarded to the requested new user in case of non-sensitiveness and also shared to rest of the user based on the sensitiveness specified by the data owner.

6 Common constructions about Outsourced data

In this section, we shall give generic constructions of CPA secure and RCCA-secure ABE systems with verifiable outsourced decryption from CPA-secure ABE system with outsourced decryption respectively. Note that our constructions can be also applied to selectively CPA secure ABE systems with outsourced decryption. And then, the resulting ABE systems with verifiable outsourced decryption are only selectively CPA-secure/RCCA-secure as well. Unlike the technique of separately encrypting an extra random message and then using this random message to commit to the true message in, the method adopted in our CPA-secure construction is encrypting a message and a random value together and then committing to the message by using the random value. Our method brings significant benefits. First, our construction is generic, and it has more compact cipher text and less computational costs. Second, such a generic CPA-secure construction can be transformed into a generic RCCA-secure construction more naturally.

6.1 Common CPA-Secure Construction

Now, we give the generic construction of CPA-secure ABE with verifiable outsourced

ISSN: 2454-9924

decryption from CPA-secure ABE with outsourced decryption. As we have said, in this construction, we employ a commitment scheme to verify the correctness of the outsourced decryption.

6.2 Moving on to Generic RCCA-Secure Construction

In this subsection, we give the generic construction of RCCA-secure ABE with verifiable outsourced decryption from CPA-secure ABE with outsourced decryption. Suppose that we shall construct an ABE scheme with verifiable outsourced decryption to work with a universe U. A set W of dummy attributes, which is disjointed from U, is used in the construction. The underlying CPA-secure ABE scheme with outsourced decryption is then required to work with universe U [W]. A dummy attribute set S W is associated to an encapsulation string com of an encapsulation scheme.

7 Conclusion

From this, Role Check Secured User Data Access with Attribute-based Encryption, Dynamic Key Generation & User Revocation System has been implemented. We have presented generic constructions of CPA-secure and RCCA-secure ABE with verifiable outsourced decryption from CPA-secure ABE with outsourced decryption, respectively. Note that the techniques involved in RCCA-secure construction can be applied in generally constructing CCA secure ABE from CPA-secure ABE. We have instantiated the CPA-secure construction in the standard model. Also of importance is the fact that a RCCA-secure ABE scheme with verifiable outsourced decryption in the standard model can be easily obtained from our generic RCCA-secure construction. We have then implemented our CPA-secure instantiation. The experimental results show that, compared with the existing selectively CPA-secure system, our instantiation has more compact cipher text and less computational costs. Additionally in future, the system enhances its performance and efficiency by reducing time consumption and cost.

8 References

[1] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, ‐Fully secure functional encryption: attribute-based encryption and (hierarchical) inner product encryption,‖ in Proceedings of the 29th Annual international conference on Theory and Applications of Cryptographic

ISSN: 2454-9924

Techniques, ser. EUROCRYPT'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 62–91.

[2] A. Sahai and B. Waters, ―Fuzzy identity-based encryption,‖ in Advances in Cryptology EUROCRYPT 2005, ser. Lecture Notes in Computer Science, R. Cramer, Ed. Springer Berlin Heidelberg, 2005, vol. 3494, pp. 457–473.

[3] B. Waters, ―Ciphertext-policy attribute-based encryption: an expressive, efficient, and provably secure realization,‖ in Proceedings of the 14th international conference on Practice and theory in public key cryptography conference on Public key cryptography, ser. PKC'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 53–70.

[4] D. Boneh, A. Sahai, and B. Waters, ―Functional encryption: Definitions and challenges,‖ in Theory of Cryptography, ser. Lecture Notes in Computer Science, Y. Ishai, Ed. Springer Berlin Heidelberg, 2011, vol. 6597, pp. 253–273.

[5] D. Boneh and J. Katz, ―Improved efficiency for CCA-secure cryptosystems built using identity-based encryption,‖ in Topics in Cryptology - CT-RSA 2005, ser. Lecture Notes in Computer Science, A. Menezes, Ed. Springer Berlin Heidelberg, 2005, vol. 3376, pp. 87–103.

[6] E. Fujisaki and T. Okamoto, ―Secure integration of asymmetric and symmetric encryption schemes,‖ in Advances in Cryptology CRYPTO '99, ser. Lecture Notes in Computer Science, M. Wiener, Ed. Springer Berlin Heidelberg, 1999, vol. 1666, pp. 537–554.

[7] G. Wang, Q. Liu, and J. Wu, ―Hierarchical attribute-based encryption for fine-grained access control in cloud storage services,‖ in Proceedings of the 17th ACM conference on Computer and communications security, ser. CCS '10. New York, NY, USA: ACM, 2010, pp. 735–737.

[8] J. Bethencourt, A. Sahai, and B. Waters, ―Ciphertext-policy attribute-based encryption,‖ in Proceedings of the 2007 IEEE Symposium on Security and Privacy, ser. SP '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 321–334

[9] J. Lai, R. Deng, C. Guan, and J. Weng, ―Attribute-based encryption with verifiable outsourced decryption,‖ IEEE Transactions on Information Forensics and Security, vol. 8, no. 8, pp. 1343–1354, Aug 2013.

[10] J. A. Akinyele, C. Garman, I. Miers, M. W. Pagano, M. Rushanan, M. Green, and A. D. Rubin, ―Charm: a framework for rapidly prototyping cryptosystems,‖ Journal of Cryptographic Engineering, vol. 3, no. 2, pp. 111–128, 2013.

[11] J. Hur and D. K. Noh, ―Attribute-based access control with efficient revocation in data outsourcing systems,‖ Parallel and Distributed Systems, IEEE Transactions on, vol. 22, no. 7, pp. 1214–1221, 2011.

ISSN: 2454-9924

[12] L. Cheung, J. A. Cooley, R. Khazan, and C. Newport, ‐Collusionresistant group key management using attribute-based encryption,‖ Group-Oriented Cryptographic Protocols, p. 23, 2007.

[13] M. Green, S. Hohenberger, and B. Waters, ‐Outsourcing the decryption of abe ciphertexts,‖ in In: Proceedings of the 20th USENIX Conference on Security, SEC 2011. San Francisco, CA, USA: USENIX Association, Berkeley, 2011.

[14] R. Canetti, O. Goldreich, and S. Halevi, ‐The random oracle methodology, revisited (preliminary version),‖ in Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, ser. STOC ’98. New York, NY, USA: ACM, 1998, pp. 209–218.

[15] S. Jahid, P. Mittal, and N. Borisov, ‐Easier: encryption-based access control in social networks with efficient revocation,‖ in Proceedings of the 6th ACM Symposium on Information, Computer andCommunications Security, ser. ASIACCS ’11. New York, NY, USA: ACM, 2011, pp. 411–415.

[16] R. Ostrovsky, A. Sahai, and B. Waters, ‐Attribute-based encryption with non monotonic access structures,‖ inProceedings of the 14th ACM Conference on Computer and Communications Security, ser. CCS ’07. New York, NY, USA: ACM, 2007, pp. 195–203.

[17] S. Muller, S. Katzenbeisser, and C. Eckert, ‐Distributed attributebased encryption,‖ in Information Security and Cryptology — ICISC 2008, P. J. Lee and J. H. Cheon, Eds. Berlin, Heidelberg: SpringerVerlag, 2009, pp. 20–36.

[18] S. Yamada, N. Attrapadung, G. Hanaoka, and N. Kunihiro, ‐Generic constructions for chosen-ciphertext secure attribute based encryption,‖ in Public Key Cryptography - PKC 2011, ser. Lecture Notes in Computer Science, D. Catalano, N. Fazio, R. Gennaro,

and A. Nicolosi, Eds. Springer Berlin Heidelberg, 2011, vol. 6571, pp. 71–89.

[19] S. Yu, C. Wang, K. Ren, and W. Lou, ‐Attribute based data sharing with attribute revocation,‖ in Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, ser. ASIACCS ’10. New York, NY, USA: ACM, 2010, pp. 261–270.

[20] T. Okamoto and K. Takashima, ‐Fully secure functional encryption with general relations from the decisional linear assumption,‖ in Proceedings of the 30th Annual Conference on Advances in Cryptology, ser. CRYPTO’10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 191–208.

[21] T. Pedersen, ‐Non-interactive and information-theoretic secure verifiable secret sharing,‖ in Advances in Cryptology – CRYPTO ’91, ser. Lecture Notes in Computer Science, J. Feigenbaum, Ed. Springer Berlin Heidelberg, 1992, vol. 576, pp. 129–140.

ISSN: 2454-9924

[22] V. Goyal, O. Pandey, A. Sahai, and B. Waters, ―Attribute-based encryption for fine-grained access control of encrypted data,‖ in Proceedings of the 13th ACM conference on Computer and communications security, ser. CCS ’06. New York, NY, USA: ACM, 2006, pp. 89–98.

ISSN: 2454-9924

Suspicious Movement Detection and Tracking based on Color Histogram

R. Thiripurasundhari #1, K. Lakshminarayanan *1
Assistant Professor *1
Mailam Engineering College, Mailam #1, *1
Sundharir93@gmail.com

Abstract - Video Monitoring systems are becoming highly important for crime investigation and the number of cameras installed in public space is increasing. However, many cameras installed at static positions are required to observe a wide and complex area. Detection of suspicious human behavior is of great practical importance. Due to changing of nature of human movements, consistent classification of suspicious human movements can be very difficult. Defining an approach to the problem of automatically finding people and detecting unusual or suspicious movements in Closed Circuit TV (CCTV) videos is our primary aim. We are introducing a system that works for monitoring systems installed in indoor environments like entrances/exits of buildings, corridors, etc. Our work presents a framework that processes video data obtained from a CCTV camera fixed at a particular location.

Keywords: Suspicious, Closed Circuit, Monitoring

I. Introduction

Investigation and so the number of surveillance cameras installed in public space is increasing. Many cameras installed at fixed positions are required to observe a wide and complex area, so observation of the video pictures by human becomes difficult. So there is a need for automation and dynamism in such surveillance systems. In order to allow the different users (operators and administrators) to monitor the system selecting different Quality of Service

ISSN: 2454-9924

(QoS) are required depending on the system status and to access live and recorded video from different localizations i.e. from their mobile devices. Automatic object detection is usually the first task in a multi-camera surveillance system and background modeling (BM) is commonly used to extract predefined information such as object's shape, geometry and etc., for further processing. Pixel-based adaptive Gaussian mixture modeling is one of the most popular algorithms for BM where object detection is

ISSN: 2454-9924

formulated as an independent pixel detection problem. It is invariant to gradually light change, slightly moving background and fluttering objects. However, it usually yields unsatisfactory foreground information (object mask) for object tracking due to sensor noise and inappropriate GM update rate, which will lead to holes, unclosed shape and inaccurate boundary of the extracted object. While sensor noise can be suppressed through appropriate filtering, it is difficult to find an optimum update rate of the model because different objects behave differently in the scene. Furthermore, important information of the object such as edge and shape are not utilized in such method. Therefore, the performance of subsequent operations such as object tracking and recognition will be degraded. In this paper, a mean shift (MS)-based segmentation is proposed for improving the object mask obtained by AGMM. By using the segmentation information, holes within the mask can be significantly reduced through in painting and better alignment between the object boundary and those of the mask can be obtained. Occlusion of moving objects is a major problem in multi-camera surveillance systems. In existing multi-camera surveillance systems,

ISSN: 2454-9924

occlusion problem is addressed by fusing the BM information obtained from the overlapped image information in adjacent cameras. These approaches, however, are not directly applicable to our non-overlapping setup. Therefore we propose to use stereo cameras, which offer additional depth information to resolve the occlusion problem. The next step in our proposed framework is object tracking over multi-camera network and it consists of two parts:

1) intra-camera tracking (tracking objects within a camera view); and 2) inter-camera tracking (associating the tracks of objects observed in different camera views). A comprehensive survey on intra-camera tracking algorithms can be found in and it can be classified into two categories in termsof tracking strategy: deterministic and probabilistic tracking. For the former, is the most popular because of its simplicity and efficacy? It only keeps a single hypothesis/candidate and utilizes the gradient of the data distribution for seeking the maximum possible candidate. Consequently, it is very computationally efficient. However, conventional MS trackeris prone to losing tracks due to rapid movement of the object. Moreover, its performance degrades considerably if

ISSN: 2454-9924

significant occlusion occurs or there are similar objects in the scene. Monitoring: Surveillance is the monitoring of the behavior, activities, or other changing information, usually of people for the purpose of influencing, managing, directing, or protecting. Surveillance is therefore an ambiguous practice, sometimes creating positive effects, at other times negative. It is sometimes done in a surreptitious manner. It most usually refers to observation of individuals or groups by government organizations, but disease surveillance, for example, is monitoring the progress of a disease in a community. The word surveillance is the French word for "watching over"; "sur" means "from above" and "veiller" means "to watch". The inverse (reciprocal) of surveillance ("to watch from below"). The word surveillance may be applied to observation from a distance by means of electronic equipment (such as CCTV cameras), or interception of electronically transmitted information (such as Internet traffic or phone calls). It may also refer to simple, relatively no- or low-technology methods such as human intelligence agents and postal interception.

II. Related Work

ISSN: 2454-9924

- The Existing methodology is a switch is attached to the door which detects any intrusion attempted by intruders.
- Image is can be stored in the server and it can be retrieve after some time
- The interrupts GSM modem and the modem sends a per-configured warning SMS to the mobile phone inthe remote location.
- Moreover there is no alert system to inform the admin when unknown object is detected.
- If the user acknowledges the pop-up, immediately a message is send back to the remote modem.

2.1 Drawback

There is no accuracy in the captured image.

- The moving object cannot be detected correctly.
- SMS alert about the motion detectionto the user.
- Image cannot be retrieve at the time of motion detection.

III. Proposed Work

ISSN: 2454-9924

In the Proposed system, the moving object is identified using the image Cauchy distribution model method. The previous frame is compared with the current frame. From that the moving object is identified. Here we can detect the exact image of the moving object. Controlling home appliances remotely with mobile applications have started becoming quite popular due to the exponential rise in use of mobile devices. Another advantage of this system is when the threshold value is reaching the limit that time server detected as a motion. Then the system will alert the user automatically by sending a GCM alert to user's mobile application. User will be using Android Mobile for the Retrieval of Images from the remote place to know whether those images are important and can be ignored.

3.1 Benefits

- High accuracy in image capturing
- Send an SMS alert to user's mobile whenever a Moving object is detected
- Image can be stored in the server and can be view at the time of motion detection.
- User can view the image, via his Android mobile itself.

IV. Absolute Effort Estimation Algorithm

It is a computational vision process of extracting foreground objects in a particular scene. A foreground object can be described as an object of attention which helps in reducing the amount of data to be processed as well as provide important information to the task under consideration. Often, the foreground object can be thought of as a coherently moving object in a scene. Absolute effort estimation algorithms a class of techniques for segmenting out objects of interest in a scene for applications such as surveillance. There are many challenges in developing a good Absolute effort estimation algorithm. First, it must be robust against changes in illumination. Second, it should avoid detecting non-stationary background objects and shadows cast by moving objects. A good background model should also react quickly to changes in

ISSN: 2454-9924

background and adapt itself to accommodate

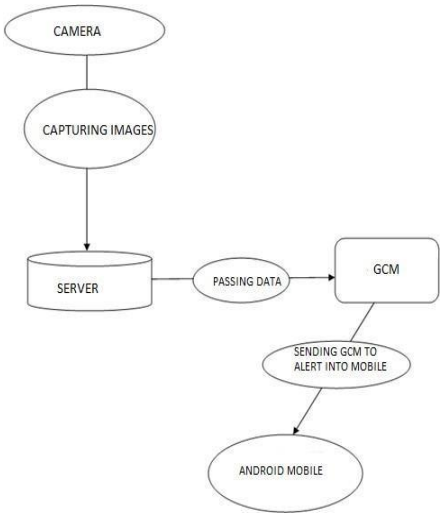
ISSN: 2454-9924

changes occurring in the background such as moving of a stationary chair from one place to another. It should also have a good foreground detection rate and the processing time for Absolute effort estimation algorithm should be real-time.

V. Cauchy Distribution Model Algorithm

The Cauchy distribution, named after Augustine Cauchy, is a continuous probability distribution. It is also known, especially among physicists, Cauchy–Lorentz distribution, Lorentz function, or Breit Wigner distribution. The simplest Cauchy distribution is called the standard Cauchy distribution. It has the distribution of a random variable that is the ratio of two independent standard normal random variables. This has the probability density function its cumulative distribution function has the shape of an arctangent function .The Cauchy distribution is often used in statistics as the canonical example of a "pathological" distribution. Both its mean and its variance are undefined. The accurate detection of the pixels at each frame is calculated by the Cauchy distribution model which uses the absolute frame differential estimation. It is expressed as follows where is the location parameter and b is the scale parameter.

VI. System Design



GCM - Google Cloud Messages

VII. Methodology

7.1 User Authentication for Application

User authentication is a means of identifying the user and verifying that the user is allowed to access some restricted service. The main aim of these modules is to authenticate the user to application to view the motion detected image this module include username and password for authentication to application the validation is based on web service in server. All subsequent server requests from a user have the session key attached to it, allowing a lookup to be made against the supplied

ISSN: 2454-9924

username to make sure the keys on the

ISSN: 2454-9924

server and device match. If the keys do not match, the server responds with an unauthorized response code. The error handling on the device causes are authorization request' to be made to the server.

7.2 GCM ID Generations for Unique ID Creation

Google-provided GCM Connection Servers take messages from a 3rd-party application server and send these messages to a GCM-enabled Android application (the "client app") running on a device. Currently Google provides connection servers for HTTP. The 3rd-Party Application Server is a component that you implement to work with your chosen GCM connection server(s). App servers send messages to a GCM connection server; the connection server en queues and stores the message, and then sends it to the device when the device is online. The Client App is a GCM-enabled Android application running on a device. To receive GCM messages,this app must register with GCM and get a registration ID. If you are using the connection server, the client app can send "upstream" messages back to the connection server. For more information on

ISSN: 2454-9924

how to implement the client app, see Implementing GCM Client.

7.3 Object Motion Detection Model

Video cameras assist in motion detection by capturing the objects of interest in the form of sets of image pixels where qualitative measurements such as recall and precision are used for assessment. The video tracker estimates the location of the object over a time by modeling the relationship between the appearance of the target and its corresponding pixel values. Determination of the relationship between an object and its image projection is very complex that makes the video tracking task difficult. Motion detection refers to the capability of the system to detect the motion and capturing the events. Motion detection is also called as activity detection, which is a software-based monitoring algorithm. It implies that when the system detects any motions the event is captured. The major application areas of motion detection methods includes visualization of traffic flow, to classify the highway lanes, driving assistance, face detection, interaction of human-machine and remote image processing.

VIII. Conclusion

In this, Suspicious Movement Detection and Tracking based on Color Histogram has

ISSN: 2454-9924

been implemented. A new object detection algorithm using color based MS separation and depth information is first implemented for improving background modeling and separation of occluded objects. The separated objects are then detected by BKF-SGM-IMS. Finally, a non-training-based object acceptance algorithm based on SP-EMD alteration measure is presented for detection of same object quoted in nearby cameras to achieve network-based detection. The usefulness of the proposed algorithms is decorated by experimental results and comparison with predictive methods.

IX. References

[1] Y. Wu, J. W. Lim, and M. H. Yang, "Online object tracking: a benchmark," in Proc. IEEE Intl. Conf. Compt. vision Pattern Recog. Jun. 2013, pp. 2411-2418.

[2] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, "Color-base probabilistic tracking," in

Proc. European Conf. Compt. Vision, 2002, pp. 661-675.

[3] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, pp. 564-577, May 2003.

[4] S. C. Chan, B. Liao, and K. Tsui, "Bayesian Kalman filtering, regularization

ISSN: 2454-9924

and compressed sampling," in Proc. IEEE Intl. Midwest Symp. on Circuits Syst. Aug. 2011, pp. 1-4.

[5] Y. C. Ho and R. C. K. Lee, "A Bayesian approach to problems in stochastic estimation and control," IEEE Trans. Automat. Contr., vol. AC-9, pp. 333-339, 1964.

[6] K. Zhang and J. T. Kwok, "Simplifying mixture models through function approximation," IEEE Trans. Neural Networks, vol. 21, pp. 644-656, Apr. 2010.

[7] I. Bilik and J. Tabrikian, "MMSE-basedfiltering in presence of non-Gaussian system and measurement noise," IEEE Trans. Aerosp. Electron. Syst. , vol. 46, pp. 1153-1170, Jul. 2010

[8] PETS2001 and PETS2009 datasets [Online]. Available: www.cvg.rdg. ac.uk/PETS2001

and

http://www.cvg.rdg.ac.uk /PETS2009.

[9] Surveillance camera network project page [Online].

Available:

http://www.eee.hku.hk/~h0995463/SCN/.

[10] Z. Zivkovic and F. V. D. Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," Pattern Recog. Letters, vol. 27, pp. 773-780, May 2006.

[11] D. Comaniciu and V. Ramesh, "Mean

ISSN: 2454-9924

shift and optimal prediction for efficient object tracking," in Proc. Int. Conf. Image Process., 2000, pp. 70-73.

[12] S. Zhang, S. C. Chan, R. D. Qiu, K. T. Ng, Y. S. Hung, and W. Liu, "On the design and implementation of a high definition multi view intelligent video surveillance system," in Proc. IEEE Intl. Conf. Signal Process. Commu. Comput., Aug. 2012, pp. 353-357.

ISSN: 2454-9924

The Graph based Query monitoring System in Data Warehousing Environment for resource allocation and redirection of workloads

Uma Pavan Kumar Kethavarapu[#], Sridevi.S.Erady [*], Govinda Raj Pandit [$]

[#]*Associate Professor, AIMS Institutions, Bangalore*

[1]*umapavanmtech@gmail.com*

[*] *Assistant Professor, IT Dept., AIMS Institutions, Bangalore*

[2]*sreedevi_erady@yahoo.com*

[$] *Associate Professor, IT Dept., AIMS Institutions, Bangalore*

[3]*grpadith@gmail.com*

***Abstract*— This paper gives the presentation about various levels of organizations ranging from small, medium, large and enterprise. The activities are common in all the levels of data warehousing environment, the only thing to consider is the amount of data and the number of systems and number of users participating in the processing of the data. The paper presents the dominating set model and balanced behaviour of the data warehousing recourses to handle the activities in better way, we are proposing a model of Query Monitor (QM) as a basic component so as to redirect and balance the workloads based on the availability of the systems in case of enterprise data the usage of big data along with Hadoop technology is presented. The importance of this work is usage of graph based model so as to get the work load estimation, creation of the resource table and redirecting the loads to the identified systems which will help the organizations and enterprises for better management of the resources.**

***Keywords*— big data, dominating set, query monitor, enterprise, hadoop, resource table.**

I. INTRODUCTION

The organizations ranging from small number of people to large are necessarily requiring the security implications so as to handle the data and activities in effective and efficient manner [1]. The ultimate goal of any organization is to serve the needs of their customers. Some organizations are planning
the transactions such as day-to-day activities in report format, some companies are maintaining the data in servers and the strategic decisions are taken from the reports generated by the server data. In some cases the data may not be available in a single location or in a single server but that may be distributed in various servers. In this case the data should be gathered from different locations and that is consolidated to form a report and further to maintain the strategic decisions. In all the above cases the security is a mandatory aspect so as to produce the correct decision by the people or company. With the usage of data warehousing tools it is possible to achieve Extraction, Transformation and Loading (ETL Process).Some of the tools to handle ETL process are Informatica, Abinitio, Data stage with parallel jobs. In data warehousing environment the reporting side we depend on Online Analytical Processing (OLAP) some of the tools are Business Objects, Cognos.These tools are used to generate the reports which are helpful for various levels of users such as end users, Business analysts, Data base designers and administrators. In data mining the pattern recognition and hidden data patterns are identified, by using data mining techniques the efficient searching is done on the data sets. Example for data mining tool is Weka.To handle all these activities the data must be secured in all aspects such as in server side, client side, networkside.A strong mechanism is required to handle the efficient security mechanism with out loss of the data from intruders and unauthorized access so as to maintain the data in secured manner. Research is done with respect to security aspects in data warehousing but lot to be done yet. The aim of the data warehousing environment is to handle strategic decisions; to achieve this data should be in secured manner new mechanisms and new methods are required for perfect management of the data, organization and users [2]. The current research focuses on invention of new metrics and mechanisms for the sake of strong secured systems. In this view we prepared some papers and the work was published in various national and international journals and we share our ideas in various national and international conferences. Further with the help of soft set computing and fuzzy systems and the latest trend in current data warehousing industry is hadoop technology and big data by using this combination of technologies we are moving to build a high-end security model for handling the security in data warehousing environment.

II. CATEGORIES OF DATA WAREHOUSING ENVIRONMENTBASED ON SCALE

The small scale industries with specified number of users such as 15 to 200 and the distribution capacity of the systems is with in the local area network then it is less complex to attempt a security measure for those kind of organizations. It is enough to maintain a single server which will handles the details of all users and projects, so the main concern of security is based on that server only. The authentication

ISSN: 2454-9924

mechanism is helpful in the handling of the security in data warehousing.

In case of medium organizations with 200 to 500 users and the distribution capacity of the systems is between Local Area Networks with some considerable complexity in the processing of the data between LANs.The best policy of maintain security for this kind of scenario is maintain security for the servers as that of small organizations along with that we need to track the data transmission with some tracing tools in such a way that if any misuse or illegal usage of the network and fraud data interpretation all these kinds of false usage need to govern so as to process the data with in medium level organizations to handle the data warehousing projects.

In case of Large organizations with 500 to 1000 capacity users and distribution capacity is wide area networks then the required security mechanisms involve the server security as that of small and medium organizations ,network tracer usage as in case of medium organization along with that to handle bulk data transfer in case of wide area networks. The usage of encryption and each time password authentication for the users who are trying to access the sensitive data in the data transfer. The categorization of users is important in large organizations to handle the data with out any false usage.

For all the above mentioned organizations the best suited method is dominating set usage[10], in this method first we need to estimate the workload of ach system participating in an activity, suppose if all the systems are common load then the work is shared equally, suppose if some of the systems are having less work to do then the work should be distributed equally to all the systems to handle this in data structures the concept AVL tree is used to balance the systems load, and it is possible to identify the system that will serve the most of the systems need such we can name as a dominating system, if such systems are available as a group it is a dominating set to process the data and to handle common activities required by the user the dominating sets are used.

In case of enterprise with number of user capacity such as greater than 1000 and distribution capacity is between various companies in different countries where they can share the same project ,and integration of the work done by different users at different locations is required. In this scenario the server security, network trace tools usage, using the encryption in the data transfer along with that one time password usage by the user while connecting with the server. The procedure is at the time of connection with the system the user need to enter the authentication to prove the identity, after that to handle the sensitive data the server will generate one password for that specific user and the same will be send to the registered users personal id, through that only the user is able to complete his task.

III. Graph Based Resource management in data warehousing

In graph theory the dominating set and reachability are the aspects where we can make use with data warehousing. We are taking one scenario of existing some N systems in a network, each system is having some capacity so as to serve the user requests. Suppose if any of the system reached to its maximum capacity of serving the user requests then how to handle the situation. Another problem we are concentrating is the dominating set where is a cluster of machines/server where we can get the entire data so as to serve the user requests. The following integrations so as to achieve the above mentioned requirements.

Case 1: Establishing a query monitor (QM), the functionality of QM is it will periodically estimate the status of workloads and how many requests are waiting for a system. Suppose any system is exceeding the number of requests allocated to its capacity then the QM will look up in the resource table which consists of the followinginformation.

System name/No	IP address	Current Load(number of hits)	Requests in queue
A	207.46.130.1	56	5
B	207.46.130.12	67	10
C	207.46.130.20	93	3
D	207.46.130.15	100	8
E	207.46.130.24	43	78

Table1: Resource Table by QM

From the above table it is clear that system D is reached to its maximum capacity. (We are assuming that number of requests processed by each system is 100) so the queue with 8 requests cannot be assigned to the system Then the QM will look up the resource table and decides that system E is having less load but it has to process the 78 requests in the queue and there is one more possibility that based on priority of the requests in the queues the QM will redirect the new requests to the system Otherwise it will go for either A or B provided the same things should be considered by the QM,the above process will be performed by the QM recursively so as to come up with one solution of subset with the possible number of systems which are chosen so as to serve the requests in the queue.

System name/No	Action Performed By QM
A	**ADJUST CURRENT QUEUE ALONG WITH SYSTEM E QUEUE**
B	**ADJUST CURRENT QUEUE ALONG WITH SYSTEM E QUEUE**
C	**ADJUST THE CURRENT QUEUE**
D	**NO ALLOCATION FOR SYSTEM D REDIRECTED TO A OR B**
E	**ALLOCATE CURRENT QUEUE AND REDIRECTS TO A,B**

Table2: Conclusion Table by QM

Procedure (Resource Monitoring)

Inputs: SYSTEMS as NODES of Graph
Current Processes as Edges of Graph

Output: Subset of systems along with allocated requests

Mechanism:

Step1: QM will traverse the Graph in BFS notation so as estimate the current workloads of the all the systems.

ISSN: 2454-9924

Step2: The result of step1 is the above table with the details of current status and queue information.

Step3: Based on the resource table the QM will decide the outcome with possible number of systems along with the assignments of new requests based on their capacity of work load.

Step4: Assigning and monitoring the status of workloads again will be done by the QM.

Case 2: The second case we are considering is the dominating set, which we are concentrating on most of the application relevant data which is residing either in a single server/cluster through which all the systems are going to acquire their needs based on the requirement. Here we are considering the case where there is a chance of reaching the dominating set itself beyond the capacity. In that case we are making use of QM conclusion to elect a system from the list as proxy to redirect a specific module data/resources to the available system with a 2-way authentication mechanism which is in the following manner.

Procedure (Redirect Workload)

Inputs: Resource graph by QM, Dominating set

Outputs: Redirect System info to User along with authentication

Mechanism:

Step1: Estimation of server capacity so as to serve the requests made by all the systems.

a) If server is able to carry on with the work load no involvement of QM.

b) else QM will redirect the requests based on the modules and number of requests to a specific module and a mapping will be established with conclusion table with allocation details.

Step3: QM will send the system information and authentication to the authorized users so as to get the final outcome to the module that user has requested.

Step4: User will have the details of system along with the authentication information.

The following figure explain the abstract view of the resource allocation in the graphs. A hypervisor is like one monitoring tool like our QM where it can have the provision of monitoring and estimating the work load of the available systems. As per the user given constraints the QM will observe the available resources along with the CPU resources and it will construct the Resource table for the current scenario as we explained in the case1.

After that the QM will take care about the dominating set and if required the redirection of requests from the dominating set is done with the QM conclusion table as described in case 2.So finally the QM will depends on resource allocation and redirection of requests.

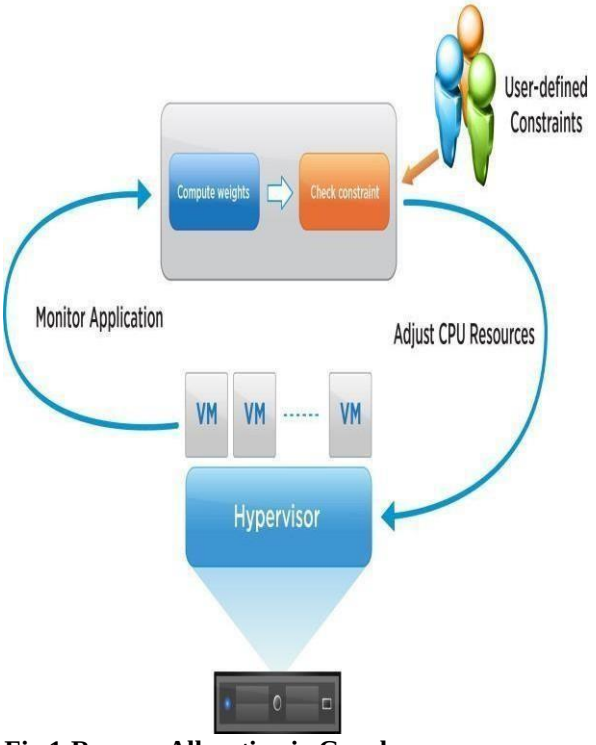


Fig 1:Resorce Allocation in Graphs

IV. FOR ENTERPRISES BIG DATA USAGE ANDHADOOP TECHNOLOGY

Big data technology enables massive data aggregation beyond what has been previously possible [11]. Given the state of today's security systems, most organizations are a long way from using these types of advanced technologies for security management. Security professionals need to get more value from the data already collected and analysed. They also need a better understanding of both current issues and impending challenges related to data. Starting with a foundational set of data management and analytic capabilities enables organizations to effectively build and scale security management as the enterprise evolves to meet Big Data challenges. When dealing with ‑Big Data,‖ the volume and types of data about IT and the business are too great to process in an ad hoc manner. Moreover, it has become increasingly difficult to secure meaningful information from the data being collected. According to the Verizon Data Breach Investigations report (2012), 91 percent of breaches led to data compromise within ‑days‖ or less, whereas 79 percent of breaches took ‑weeks‖ or more to discover. The following are major needs of big data usage

- ‑Scaling out‖ rather than ‑scaling up‖, since centralizing all this data will be practically impossible.
- Analytics and visualization tools that support security analyst specialties. Security professionals require specialized analytic tools to support their work.

ISSN: 2454-9924

- Network forensics analysts need full reconstruction of all log and network information about a session to determine precisely what happened.

Threat intelligence to apply data analytic techniques to the information collected. Organizations require a view of the current external threat environment in order to correlate with information gathered from within the organization itself. This correlation is key for analysts to gain a clear understanding of current threat indicators and what to look for.

Security organizations today need to take a ‑Big Data‖ approach. Eliminate tedious manual tasks in routine response or assessment activities.

- Use business context to point analysts toward highest impact issues. Security teams need to be able to map the systems they monitor and manage back to the critical applications and business processes they support.
- Present only the most relevant data to analysts. Security professionals often refer to ‑reducing false positives.‖
- See ‗over the horizon.‘ Defence against modern threats is a race against time. The system needs to provide early warning –and eventually predictive model
- Start by implementing a security data infrastructure that can grow with you. This involves implementing an architecture that can not only collect detailed information about logs, network sessions, vulnerabilities, configurations, and identities, but also human intelligence about what systems do and how they work.
- Deploy basic analytic tools to automate repetitive human interactions.
- Create visualizations and outputs that support major security functions. Some analysts will only need to see the most suspicious events with some supporting detail. Malware analysts will need a prioritized list of suspect files and the reasons why they are suspect.

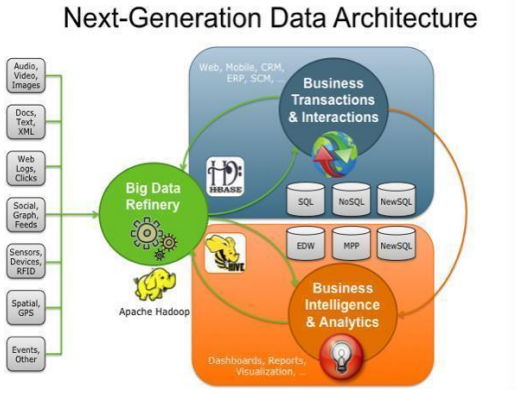


Fig. 2 Big Data Architecture

A. *Hadoop Basics*

A software framework that supports distributed computing using Map Reduce [12]

- Distributed, redundant file system (HDFS)

Job distribution, balancing, recovery, scheduler, etc.

- **Map Reduce:** A programming paradigm that is composed of two functions (~ relations)

Map Reduce

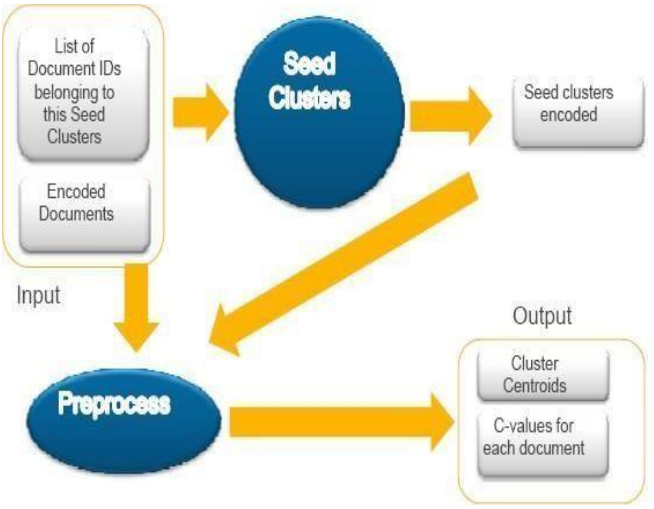


Fig. 3 Processing of Data in Hadoop

The Security of Big Data Infrastructure considers the following principles

- Evaluate the security capabilities of big data infrastructure
- Do the available tools provide needed security features?
- What security models can be used when implementing big data infrastructure?
- . Identify techniques to enhance security in big data frameworks (e.g., data tagging approaches, sHadoop)
- Conduct experiments on enhanced security framework implementations

V. CONCLUSION AND FUTURE WORK

The overall concept we discussed in this paper is starting from the small organization to enterprise the activities of data warehousing are common that is OLTP, ETL and OLAP.But the differentiation is regarding with data, whether the warehousing uses parallel and distributed environments. Depending on the level of organization the data handling mechanism will be changed.Upto the large organizations according to our view the usage of dominating sets with balancing nature of the load of

ISSN: 2454-9924

the system is better, where as in case of enterprises we need to move for big data analytics another important criterion is the security implementation depending on number of users and kind of activity taken in the data warehousing environment. Some security measurements and requirements are specified in the paper. The future scope and enhancement for this work is constructing a common security mechanism starting from requirements gathering up to maintenance phase, and more over irrespective of the kind of data warehousing the common security mechanism is required to implement.

REFERENCES

[1] Dr..S.L Gupta,Sonali Mathur,Palal Modi, Data Warehouse Vulnerability and Security, International Journal Of Scientific & Engineering Research Volume 3,Issue5,May-2012.

[2] J. Mohamed Salah Gouider,Amine Farahat,Building Data Warehouse National Social Security Fund Of The Republic Of Tunisia, International Journal Of Database Management Systems(IJDMS),Vol 2,No2,May 2010.

[3] S. V.M.NavaneethaKumar, Dr.C.Chandrasekhar, Security Of data Warehousing Server, International Journal Of Computer Applications, Vol-III, No.4, Dec 2010.

[4] M. Robert Winter,Olivera Marjanovic,Barbara H.Wixom,Introduction to the Business Intelligence and Data Warehousing Minitrack,45 th Hawaii International Conference On System Sciences,IEEE-2012.

[5] R. Shashank Saroop,Manoj Kumar, Comparison Of Data Warehouse Design Approaches From User Requirement to Conceptual Model:A Survey, International Conference On Communication System and Network Technologies,IEEE-2011.

[6] Stefano Rizzi,Albrt Abello,Jens Lechtenborger,Juan Trujillo, Research in Data Warehouse Modeling and Design: Dead or Alive?ACM Transactions NOV-2012.

[7] M. Veronique Limere,Aditya Pradhan,Melih Ceilk,Mallory Soldner,Warehousing Efficiency In a Small Warehouse,IEEE 2011.

[8] *Marcel Danilescu,Data Security management applying trust policies for small organizations,ad hoc organizations and Virtual OrganizationsJournal of Accounting and management,Vol 2,no.3-2012*.

[9] -Xuejian Yan,Xueqing Li,A Multidimensional Data Analysis Based on MDA for Educational Data Warehousing, The 6th International Conference on Computer Science and Education,IEEE,Aug-2011.

[10] Uma Pavan Kumar Kethavarapu,Dr.S.Saraswathi,,The Requirements of Parallel Data Warehousing Environment to Improve the Performance with Dominating sets for Next Generation Users.Intenational Journal Of Computer Science and Information Security,Vol.10,No.5,May 2012.

[11] Sam Curry, Engin Kirda, Eddie Schwartz, William H.Stewart,Amit Yoran,Big data Fuels Intelligence-Driven Security, January 2013.

[12] HDFS Architecture Guide, White Paper Apache Software Foundation 2012.

The Author Named **Uma Pavan Kumar Kethavarapu** he received his **M.Tech** from NIET, Sattenapalli affiliated to **JNTUK**, Kakinada, He is having total of 9 years of teaching experience in various levels such as lecturer, Assistant professor and Associate Professor. His research interest are **Data warehousing, Data bases, distributed and parallel systems. He is having papers published in national, international journals and attended various national and international conferences.**

The Author Named **Sreedevi.S.erady** she received her MCA from DOEACC centre, affiliated to Calicut University, Kerala, she is having total of 7 years of teaching experience in various levels such as lecturer, Assistant professor. The author is SET qualified from Karnataka. Her research interest are Data **warehousing, Data mining, and Security aspects in data bases and data warehousing, Distributed environments.**

The author named **Govinda raj pandit** MSC CS from University of Mysore, he is having total of 13 years of teaching experience. His research interests are **Network security, data warehousing, algorithms, data structures and operating systems.**

ISSN: 2454-9924

User Activity Monitoring for Dynamic and Flexible Group Key Generation

R. Shanmuga Sundaram #1, T. Priya Rathika Devi *1
Mailam Engineering College, Mailam #1, *1
Shanmugasundaram2008@gmail.com

Abstract -In cloud user data can be easily shared and viewed across the shared medium. To make sure the shared data can be verified publicly, users in the group need to calculate signatures on all the blocks in shared data. Various blocks in shared data are generally determined by different users due to data modifications indicated by different users. In the previous methods, there is no Security in the Cloud is enforced. In the proposed method**,** Data Owner updates the information to the Remote Cloud Server for Data Access. Data owner appoints Members for Data Utility and Data updating. Members have to get permission for the Data updating from the Data Owner. Members will have their User Name, Key, and Group Key for Access. If Existing member is removed from that Group, Group Key is automatically changed and updated to all the Members of that Group. The modified work is Group Key can be changed in case of New Member is added in that Group or Existing Member is Resigned by themselves from the Group or Data Owner Terminates the Member or Cloud Terminates the Member in case of Misbehavior (DDOS Attack, Same Data Download), updated new key is sent to the corresponding users through Email.

Keywords: Data owner, Signatures, Cloud server, Permission, Group key

1Introduction

With data storage and sharing services (such as Drop box and Google Drive) provided by the cloud, people can easily work together as a group by sharing data with each other. More specifically, once a user creates shared data in the cloud, every user in the group is

ISSN: 2454-9924

able to not only access and modify shared data, but also share the latest version of the shared data with the rest of the group. Although cloud providers promise a more secure and reliable environment to the users,the integrity of data in the cloud may still be compromised, due to the existence of

ISSN: 2454-9924

hardware/ software failures and human errors. To protect the integrity of data in the cloud, a number of mechanisms have been proposed. In these mechanisms, a signature is attached to each block in data, and the integrity of data relies on the correctness of all the signatures. One of the most significant and common features of these mechanisms is to allow a public verifier to efficiently check data integrity in the cloud without downloading the entire data, referred to as public auditing (or denoted as Provable Data Possession). This public verifier could be a client who would like to utilize cloud data for particular purposes (e.g., search, computation, data mining, etc.) or a third-party auditor (TPA) who is able to provide verification services on data integrity to users. Most of the previous works focus on auditing the integrity of personal data. Different from these works, several recent works focus on how to preserve identity privacy from public verifiers when auditing the integrity of shared data. Unfortunately, none of the above mechanisms considers the efficiency of user revocation when auditing the correctness of shared data in the cloud. With shared data, once a user modifies a block, she also needs to compute a new signature

ISSN: 2454-9924

for the modified block. Due to themodifications from different users, differentblocks are signed by different users. Forsecurity reasons, when a user leaves thegroup or misbehaves, this user must berevoked from the group. As a result, thisrevoked user should no longer be able toaccess and modify shared data, and thesignatures generated by this revoked user areno longer valid to the group. Therefore,although the content of shared data is notchanged during user revocation, the blocks,which were previously signed bytherevoked user, still need to be re-signed by anexisting user in the group. As a result, the integrity of the entire data can still be verified with the public keys of existing users only. Since shared data is outsourcedto the cloud and users no longer store it onlocal devices, a straightforward method tore-compute these signatures during userrevocation is to ask an existing user (i.e.,Alice) to first download the blocks previously signed by the revoked user (i.e.,Bob), verify the correctness of these blocks.

2 Related works

In the existing system, there is no Security in the Cloud is enforced. Here it provides some drawbacks they are; provide less

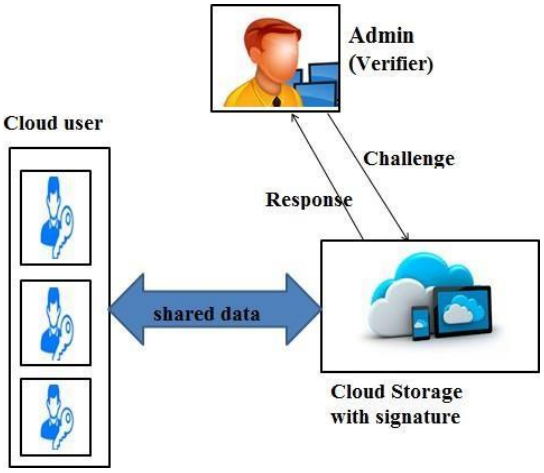
ISSN: 2454-9924

security and poor data integrity and confidentiality

2.1 Proposed work

Data Owner updates the information to the Remote Cloud Server for Data Access. Data owner appoints Members of Data Utility and Data updating. Members have to get permission for the Data updating from the Data Owner. Members will have their User Name, Key, and Group Key for Access. Either If Existing member is removed from that Group, Group Key is automatically changed and updated to all the Members of that Group. Group Key can be changed in case of New Member is added in that Group also. Member can resign from the Group by themselves or Data Owner can terminate the Member or can be Cloud Terminates the Member in case of Misbehavior (DDOS Attack, Same Data Download). Finally the proposed system provides less efficiency and high performance.

3 System Model



We assume the cloud itself is semi- trusted, which means it follows protocols and does not pollute data integrity actively as a malicious adversary, but it may lie to verifiers about the incorrectness of shared data in order to save the reputation of its data services and avoid losing money on its data services. In addition, we also assume there is no collusion between the cloud and any user during the design of our mechanism. Generally, the incorrectness of share data under the above semi-trusted model can be introduced by hardware/software failures or human errors happened in the cloud. Considering these factors, users do not fully trust the cloud with the integrity of shared data. To protect the integrity of shared data, each block in shared data is attached with a signature,

ISSN: 2454-9924

which is computed by one of the users in the

ISSN: 2454-9924

group. Specifically, when shared data is initially created by the original user in the cloud, all the signatures on shared data are computed by the original user. After that, once a user modifies a block, this user also needs to sign the modified block with his/her own private key. By sharing data among a group of users, different blocks may be signed by different users due to modifications from different users. When a user in the group leaves or misbehaves, the group needs to revoke this user. Generally, as the creator of shared data, the original user acts as the group manager and is able to revoke users on behalf of the group. Once a user is revoked, the signatures computed by this revoked User become invalid to the group, and the blocks that were previously signed by this revoked user should be re-signed by an existing user's private key, so that the correctness of the entire data can still be verified with the public keys of existing users only.

4 Methodologies

Network Construction

This module is developed in order to create a dynamic network. In a network, nodes are interconnected with a particular group and the information can be shared among them.

ISSN: 2454-9924

For the successful data transfer the network must be properly controlled and handled. Every node is interconnected & this forms a network.

4.1 Server in Cloud

Here the server will have the entire details about all the group information. It distributes the data to client in a particular group. Server is responsible for maintaining all the group information. If any user will removed from a particular group means it will instruct to the group member to change the group id and send the SMS to all group members.

4.2 Cloud User Status

Users can be moved from one group to another group and he will also participated in more than one group. Depending upon the user status they can share their information with the group member. All the user status information is to be maintained here. If a new user login or the existing user logged inor logged out all that information about a user must maintained for authenticate.

4.3 Generation of Group Key

In this module we have to create group key as well as the individual key then share the key between the group members via SMS.

ISSN: 2454-9924

Any changes occurs in the group then change the group key and then send that key to the other entire group member. This process is done whenever any changes made in that group.

4.4 Cloud Data Access

If a user wants to access any information about any user then he will give his individual key as well as the group key. If he want to access the information about the user, but the user is not belongs to their group is not possible. He can only access the user's information within their group only. Without knowledge of the other group key it is not possible to access the information.

5 PANDA

Based on the new proxy re-signature scheme and its properties in the previous section, we now present Panda—a public auditing mechanism for shared data with efficient user revocation. In our mechanism, the original user acts as the group manager, who is able to revoke users from the group when it is necessary. Meanwhile, we allow the cloud to perform as the semi-trusted proxy and translate signatures for users in the group with re-signing keys. As emphasized in recent work, for security reasons, it is

ISSN: 2454-9924

necessary for the cloud service providers to storage data and keys separately on differentservers inside the cloud in practice. Therefore, in our mechanism, we assume the cloud has a server to store shared data, and has another server to manage re-signing keys. To ensure the privacy of cloud shared data at the same time, additional mechanisms, such as, can be utilized. The details of preserving data privacy are out of scope of this paper. The main focus of this paper is to audit the integrity of cloud shareddata.

6 Literature Review

[2] Describes about, with data services in thecloud, users can easily modify and share data as a group. To ensure data integrity can be audited publicly, users need to compute signatures on all the blocks in shared data. Different blocks are signed by different users due to data modifications performed by different users. For security reasons, once a user is revoked from the group, the blocks, which were previously signed by this revoked user, must be re-signed by an existing user. The straightforward method, which allows an existing user to download the corresponding part of shared data and re-sign it during user revocation, is inefficient

ISSN: 2454-9924

due to the large size of shared data in the cloud. In this paper, we propose a novel public auditing mechanism for the integrity of shared data with efficient user revocation in mind. By utilizing proxy re-signatures, we allow the cloud to re-sign blocks on behalf of existing users during user revocation, so that existing users do not need to download and re-sign blocks by themselves. In addition, a public verifier is always able to audit the integrity of shared data without retrieving the entire data from the cloud, even if some part of shared data has been re-signed by the cloud. Experimental results show that our mechanism can significantly improve the efficiency of user revocation.

[14] Describes about**,** in a proof-of-irretrievability system, a data storage center must prove to a varied that he is actually storing all of a client's data. The central challenge is to build systems that are both efficient and provably secure that is, it should be possible to extract the client's data from any prove that passes a verification check. In this paper, we give the first proof-of-irretrievability schemes with full proofs of security against arbitrary adversaries in the strongest model, that of Juels and Kaliski. Our scheme, built from BLS signatures and secure in the random oracle

ISSN: 2454-9924

model, has the shortest query and response of any proof-of-irretrievability with public variability. Our second scheme, which builds elegantly on pseudorandom functions (PRFs) and is secure in the standard model, has the shortest response of any proof-of- irretrievability scheme with privatevariability (but a longer query). Both schemes rely on holomorphic properties to aggregate a proof into one small authenticator value. [10] Describes about, Cloud Computing has been envisioned as the next generation architecture of IT Enterprise. In contrast to traditional solutions, where the IT services are under proper physical, logical and personnel controls, Cloud Computing moves the application software and databases to the large data centers, where the management of the data and services may not be fully trustworthy. This unique attribute, however, poses many new security challenges which have not been well understood. In this article, we focus on cloud data storage security, which has always been an important aspect of quality of service. To ensure the correctness of users' data in the cloud, we propose an effective and flexible distributed scheme with two salient features, opposing to its predecessors. By utilizing the

ISSN: 2454-9924

homomorphism token with distributed verification of erasure-coded data, our scheme achieves the integration of storage correctness insurance and data error localization, i.e., the identification of misbehaving server(s). Unlike most prior works, the new scheme further supports secure and efficient dynamic operations on data blocks, including: data update, delete and append. Extensive security and performance analysis shows that the proposed scheme is highly efficient and resilient against Byzantine failure, malicious data modification attack, and even server colluding attacks. [11] describes about, We introduce a model for provable data possession (PDP) that allows a client that has stored data at an untrusted server to verify that the server possesses the original data without retrieving it. The model generates probabilistic proofs of possession by sampling random sets of blocks from the server, which drastically reduces I/O costs. The client maintains a constant amount of metadata to verify the proof. The challenge/response protocol transmits a small, constant amount of data, which minimizes network communication. Thus, the PDP model for remote data checking supports large data sets in widely-distributed

ISSN: 2454-9924

storage systems. We present two provably-secure PDP schemes that are more efficient than previous solutions, even when compared with schemes that achieve weaker guarantees. In particular, the overhead at the server is low (or even constant), as opposed to linear in the size of the data. Experiments using our implementation verify the practicality of PDP and reveal that the performance of PDP is bounded by disk I/O and not by cryptographic computation.

7 Conclusion

From this, User Activity Monitoring for Dynamic and Flexible Group Key Generation has been implemented. We have proposed a new public auditing mechanism for shared data with efficient user revocationin the cloud. When a user in the group is revoked, we allow the semi-trusted cloud to re-sign blocks that were signed by the revoked user with proxy re-signatures. Experimental results show that the cloud can improve the efficiency of user revocation, and existing users in the group can save a significant amount of computation and communication resources during user revocation. In future, our system will enhance the efficiency also provides securityin cloud.

8 References

[1] A. Shamir, “How to Share a Secret,” Comm. ACM, vol. 22, no. 11, pp. 612-613, Nov. 1979.

ISSN: 2454-9924

[2] Boyang Wang,Public Auditing for Shared Data with Efficient User Revocation in the Cloud

[3] B. Wang, B. Li, and H. Li, "Public Auditing for Shared Data with Efficient User Revocation in the Cloud," Proc. IEEE INFOCOM, pp. 2904-2912, 2013.

[3] B. Wang, S.S.M. Chow, M. Li, and H. Li, "Storing Shared Data on the Cloud via Security-Mediator," Proc. IEEE 33rd Int'l Conf. Distributed Computing Systems (ICDCS'13), pp. 124-133, July 2013.

[4] B. Wang, B. Li, and H. Li, "Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud," Proc. IEEE CLOUD, pp. 295-302, 2012.

[5] C. Wang, Q. Wang, K. Ren, and W. Lou, "Ensuring Data Storage Security in Cloud Computing," Proc. 17th ACM/IEEE Int'l Workshop Quality of Service (IWQoS'09), pp. 1-9, 2009.

[6] C. Wang, Q. Wang, K. Ren, and W. Lou, "Towards Secure and Dependable Storage Services in Cloud Computing," IEEE Trans. Services Computing, vol. 5, no. 2, pp. 220-232, Jan. 2012.

[7] B. Wang, H. Li, and M. Li, "Privacy-Preserving Public Auditing for Shared Cloud Data Supporting Group Dynamics," Proc. IEEE Int'l Conf. Comm. (ICC'13), pp. 1946-1950, June 2013.

[8] B. Wang, B. Li, and H. Li, "Knox: Privacy-Preserving Auditing for Shared Data with Large Groups in the Cloud," Proc. 10th Int'l Conf. Applied Cryptography and Network Security (ACNS'12), pp. 507-525, June 2012.

ISSN: 2454-9924

559, Oct.Dec. 2013.

[9] C. Wang, Q. Wang, K. Ren, and W. Lou, “Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing,” Proc. IEEE INFOCOM, pp. 525-533, 2010.

[10] Cong Wang, Ensuring Data Storage Security in Cloud Computing

[11]Giuseppe Ateniese, Provable DataPossession at Untrusted Stores

[12] G. Ateniese, R. Burns, R. Curtmola, J.Herring, L. Kissner, Z. Peterson, and D. Song, “Provable Data Possession atUntrusted Stores,” Proc. 14th ACM Conf.Computer and Comm. Security (CCS’07), pp. 598-610, 2007.

[13] G. Ateniese and S. Hohenberger, “Proxy Re-signatures: New Definitions, Algorithms and Applications,” Proc. 12th ACM Conf. Computer and Comm. Security(CCS’05), pp. 310-319, 2005.

[14] Hovav Shacham , Compact Proofs

[15] H. Shacham and B. Waters, “CompactProofs of Retrievability,” Proc. 14th Int’lConf. Theory and Application of Cryptologyand Information Security: Advances in Cryptology (ASIACRYPT’08), pp. 90107, 2008.

[16] H. Wang, “Proxy Provable Data Possession in Public Clouds,” IEEE Trans. Services Computing, vol. 6, no. 4, pp. 551-559, Oct.Dec. 2013.

[13] H. Wang, “Proxy Provable Data Possession in Public Clouds,” IEEE Trans. Services Computing, vol. 6, no. 4, pp. 551-

ISSN: 2454-9924

[17] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Kon-winski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp. 50-58, Apr. 2010.

[18] M. Blaze, G. Bleumer, and M. Strauss, "Divertible Protocols and Atomic Proxy Cryptography," Proc. Int'l Conf. the Theory and Application of Cryptographic Techniques (EUROCRYPT'98), pp. 127-144, 1998.

[19] N. Cao, S. Yu, Z. Yang, W. Lou, and Y.T. Hou, "LT Codes-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM, pp. 693-701, 2012.

[20] M. Li, N. Cao, S. Yu, and W. Lou, "FindU: Private-Preserving Personal Profile Matching in Mobile Social Networks," Proc. IEEE INFOCOM, pp. 2435-2443, 2011.