

Communication Management of Large Projects in Big Data Environment

Suhail Memon, Wang Changfeng, Shahid Rasheed, Zulfiqar Hussain Pathan,
QiuYixin and Liu Yanping

*School of Economics and Management, Beijing University of Posts and
Telecommunications, China*

*Suhail_sam06@yahoo.com; wangcf@bupt.edu.cn; shahidrasheed@outlook.com;
zulfi2k3bcs@yahoo.com; 644716494@qq.com; 1162272803@qq.com*

Abstract

The purpose of this paper is to provide a model for communication in large scale projects in Big Data environment. Given the fact that Big Data is going to be part of every business organization very soon, this study aims to present a new communication management strategy in order to accommodate the change in technology of data base storage systems. The complex large projects always emerge with different modes and have different attributes for processing the data. In this model different data sources of a large scale project are connected to the enterprise data ware house using Information Inference Framework (IIF). This new approach may provide an enhanced way of communication management for activities in large scale projects.

Keywords: *Communication Management, Large Scale Projects, Big Data Environment, Information Inference Framework (IIF)*

1. Introduction

Large scale projects emerge with different modes and in different situations. The large scale projects and programs granted and commissioned by governments have multiple enterprise partners, and are usually complex, uncertain, and risky [1, 2]. Costs of these projects usually range from \$100 million to as much as several billion dollars. Different literatures present different definitions of large projects. Merrow has defined that very large scale projects are worth \$ 500 million or more [3]. Similarly, another definition outlines that the large scientific projects are usually over \$ 1 billion, have many international partners, and may span over decades [4]. Moreover, the large scale infrastructure projects are said to have higher costs than the large scale research projects [3]. Although some studies tend to differentiate large projects from programs [5], this reading conceives them in similar meanings. In mega projects continuous reporting and communication methods are highly important to manage large sets of data. It is because the companies and organizations involved in these endeavours need frequent communication with other stakeholders in order to keep abreast with the information updates. Some researchers have argued that many large scale projects fail [6], while some others maintain that more than 50 % of management problems are caused by poor communication [7].

Since the public and private stakeholders keep looking for more efficient methods to enhance the quality of communication, Big Data has enticed greater attention all around the world [8]. To improve the productivity and growth in large projects, each stakeholder can use certain features of Big Data so that data can be transformed into products and services [9]. Large scale projects accompany larger and complex systems of information linked through interfaces; these communication interface systems produce bigger amounts

of data sets that are complex, and therefore need efficient processing and storage for data manipulation [10].

Moreover, the domains of large scale projects are expanding by time. As the projects are spreading their length geographically, the data and information are also growing in means of size and processing. The rising rates of complication in large scale projects [11] have also obligated the organizations development and adopt more robust infrastructures and architectures to handle the database systems. Either the irrelevance of data or the unavailability of information may consequence the projects towards failure [12]. So, the understanding of projects complexity needs specific information including data processing of knowledge for decision making [13]. Growing organizations are using advanced computation techniques [14]. In such techniques they are exploiting sophisticated data processing models to best handle the growing complexity of data in systems.

Large scale machines similar to Hadron Collider [15] produce zeta bytes of data, share it with several stakeholders, and process it across hundreds of nodes that are geographically apart. Large scale collaboration depends upon the functioning of distributed architecture. More prominent tools and higher network architecture with fast access are therefore becoming backbone of large scale projects. The increasing size of data also brings in the need for end-to-end network framework [16]. These frameworks embrace technologies like MPLS and SONET that can be used for communication management of large scale projects in Big Data environment.

This paper is going to represent a model of communication management for such mega projects. The purpose of this model is to connect different data sources of a large project such as Finance, Sales, Human Capital Management (HCM), Procurement, and Customer Relationship Management (CRM) *etc.*, with the enterprise data warehouse. With the help of Information Inference Framework (IIF), this model infers different types of data and sends back to the enterprise data ware house using Application Program Interfaces (APIs) where it is accessed by different external clients and stakeholders.

Rest of the paper is arranges as follows. Section 2 outlines the significance of communication amanagement in projects, while the concept and implications of Big Data are delineated in Section 2. In Section 4 this study proposes a communication management model for large scale projects in Big Data environment and presents its details. Finally Section 5 concludes the overall work.

2. Communication Management and its Significance for Projects

Communication has got strong influence in advanced technology development projects. It may be regarded as an interface between different cells of a large scale project. The relationship between stakeholders and success of a project highly depend upon the communication [17]. The needs of communication management inside project management discipline are signified by the world's leading project management bodies in the form of project management standards. PMOBK guide issued by Project Management Institute (PMI) (U.S.A.) has dedicated a special chapter on this knowledge area [18]. Similarly the Managing Successful Programs (MSP) guide by the Office of Government Commerce (OGC) U.K. has classified it as a designated function for projects and programs management [19]. Data management systems that are used in organizations have got major influence on communication. In addition to information processing, the source of information to complete the transaction and integration is really important in projects [20].

In the communication environment, electronic medium and data management has gained a unique significance. The project stakeholders including employees and members can exchange pertinent information through emails, messaging, paging, data repositories, and many other communication technologies. Big data also describes communication

management by extracting of value from modern data sources such as public information, cellular phones, car enable sensors, and other technological advents of present and future. A project supplies various sets of information to stakeholdersthat are important for suppliers, procurement, and logistic functions *etc.* Therefore communication and project performance bears a greater correlation where big and complex projects need added efficiency in communication systems [21]. Most of the information in high technology business environment is exchanged through web based programs or integrated enterprise resource programs. Corporate communication is a good example of using this web based and integrated enterprise resource programfor business communication and organizational communication relation [22]. Communication plays the role of glue to integrate internal and external elements of an organization together [23]. Following Figure (Figure 1) explains different data sources and their interfacing with an Enterprise Data Warehouse.

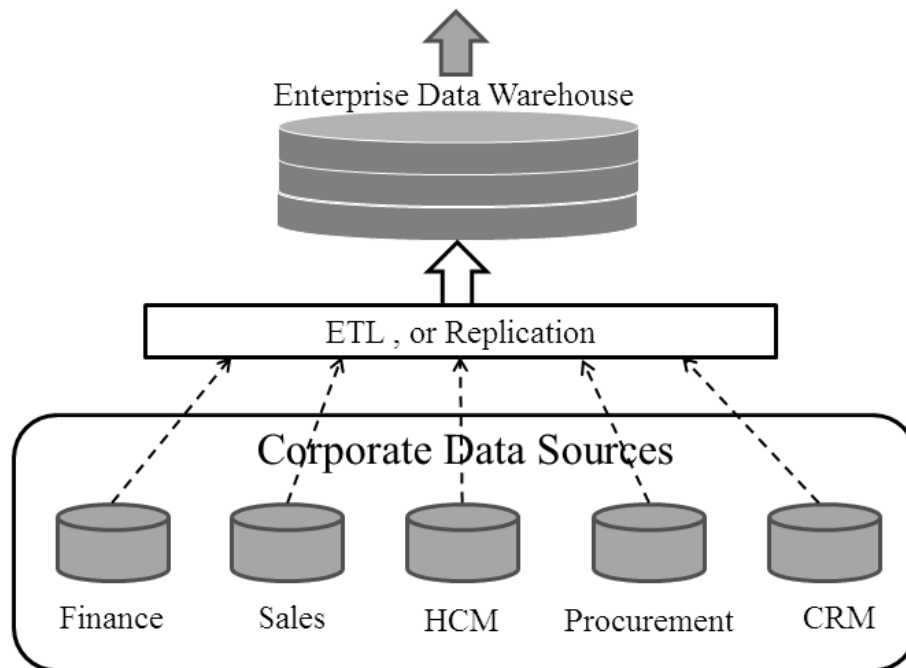


Figure 1. Enterprise Data Warehouse

3. Big Data and its Implications

Big Data explains the information management policy which combines different kinds of traditional and modern data sources. Big data introduced two popular technologies for processing and storage, which are No SQL Database and Apache Hadoop. Big data contains huge data sets that are continuously growing over time. According to some studies, the capacity of data in 2011 was 1.8 zeta bytes [24]. By year 2020 it is expected to grow by almost 50 times and will require much bigger network infrastructures to handle it. The large scale Big Data also shows special trends in mobility, faster access, and consumption of data [25]. The significance in data analytics by the Radoop which consists of Rapid Miner and Hadoop has got remarkable attention. Another main dimension of Big Data is cloud computing which is a framework for climate of various scales [26].

The network architecture of Big Data is the backbone of current collaborative infrastructures, where geographically scattered partners can interact for variety of reasons. It should therefore be efficient and smart. It has become very necessary for enterprises to mine the data to stay in the digital market. The most important competitive advantage for an organization comes from the apt analysis of data, because it produces

wealth of information. However, the volumes of Big Data sets are very ambiguous to manage. This ambiguity materializes because different type of data is extracted from different types of resources. Big data domains primarily consist of two data formations, unstructured and structured. As opposed to the unstructured data, the structured data sets come with a predefined syntax. The structured data is therefore relatively easier to process than the unstructured. No SQL Database and Apache Hadoop are capable to process both the said types of data. Big data technical tools are capable to run ad-hoc queries which can access large data sets very fast and may help the organizations to develop businesses faster including an increase in their customers and services base. Following Figure (Figure 2) describes the arrangement of Big Data at a conceptual level.

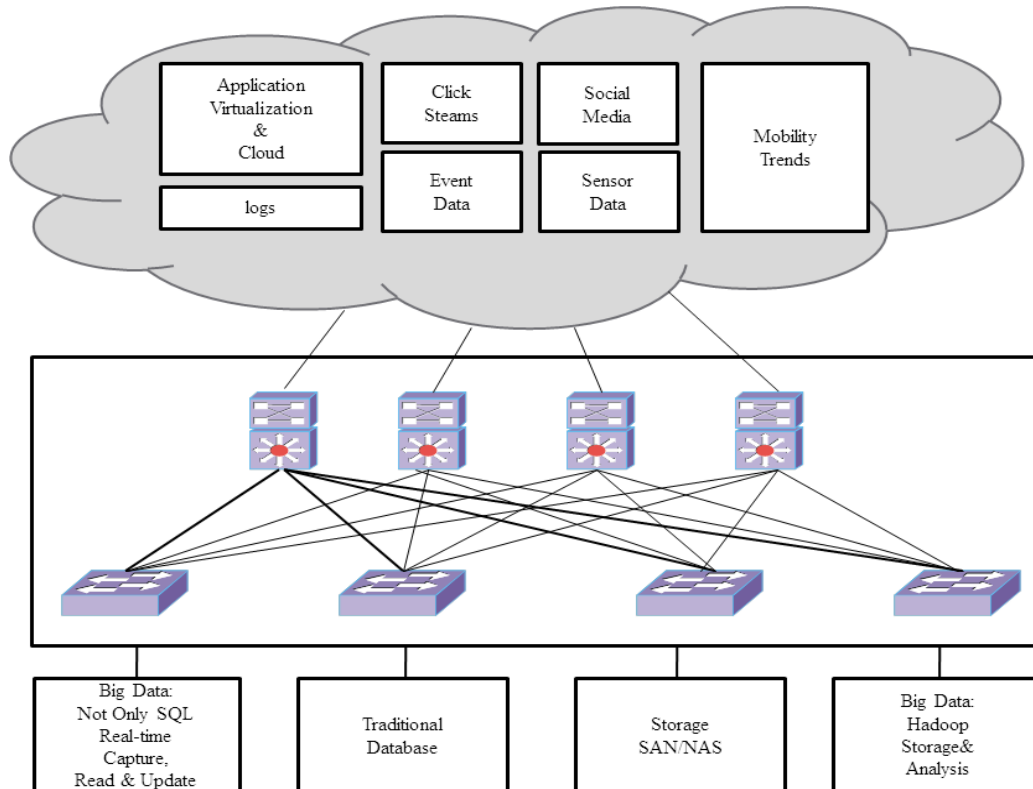


Figure 2. The Big Data Concept

4. Proposed Model for Communication Management of Large Scale Projects in Big Data Environment using IIF Framework

Figure 3 below sketches the nodes and flows of the proposed communication model. This model consists of a graph shape of connecting different enterprise data ware houses of metadata, having data sources such as finance, sales, HCM, procurement, and CRM *etc.* It helps the enterprise stakeholders in disseminating and exploring the data. This model especially provides mechanism for aggregating different metadata. The model is designed to embrace the Information Inference Framework (IIF) for large scale projects accessing Big Data. The infrastructure represents architecture for describing and executing different data flows for large data sets. For processing the data flow in a distributed Environment we use Apache Hadoop which gives access application via computer cluster. The IIF is an open source software by Rapid Miner [27] for data analytics. This tool uses a friendly graphical user interface for creating data flows from different metadata nodes. Extract, Transfer, and Load (ETL) data flow nodes are defined in the library of IIF. Rapid Miner has an advantage to carry a design that contains the

ability to process larger data sets. In order to plug-in, Rapid Miner uses Radoop for processing huge data.

Compared to other models being used in Big Data environments, this communication management model has a distinction to use IIF. The other models apply their own data processing tools using distinct service for processing a module, while this approach uses a cluster application attached with single processing module that may use Hadoop framework to communicate with other modules. In addition to Hadoop, GATE and UIMA approaches can also be used for extraction of information. In communication management of large data processing system in Big Data, IIF provides prominent tools for determining data processing system. It consists of data flow nodes, data stores for communication of data, and execution tools for data flow. IIF framework is applicable on top domain of Apache Hadoop for distributed computation of big sizes of data. For integration of new nodes in a cluster, Apache Hadoop stores data files in Hadoop Distributed File System (HDFS). On the other hand Map Reduce is used to write the logic of data flow nodes for execution on the cluster.

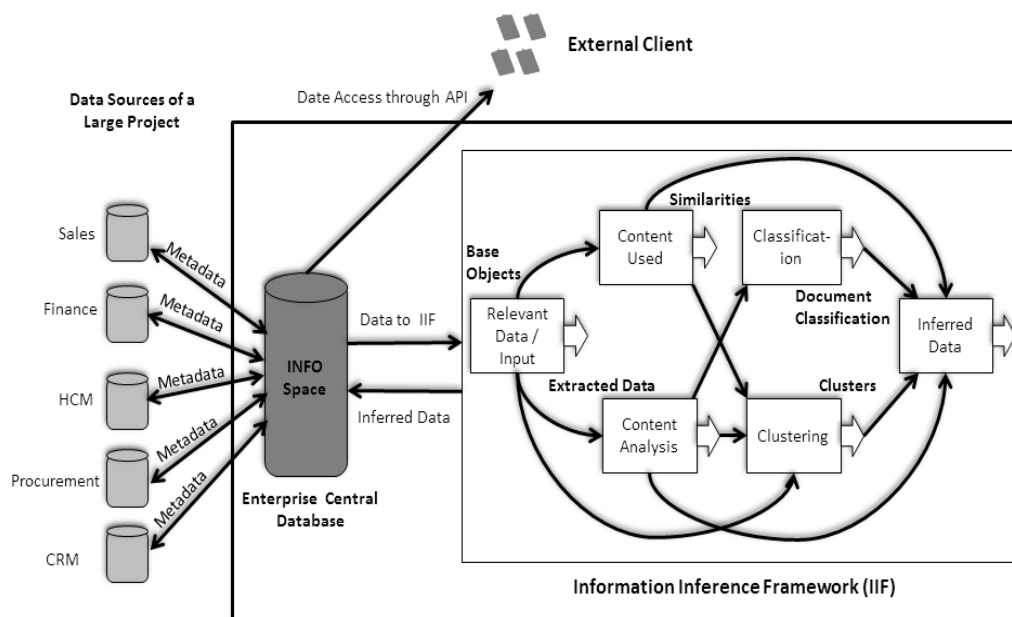


Figure 3. Model of Communication Management of Large Projects in Big Data Environment using IIF

This model contains Information Inference System Framework which is based on D-NET software toolkit [28]. The functionality of this software toolkit is to define and execute Meta data. In this model of communication management, the architecture is designed to collect XML serials of different Meta data from different domains *e.g.*, Finance, Sales, HCM, Procurement and CRM, which are different data sources in large projects as shown in the above figure. These XML serials are converted to become objects, and stored in enterprise central database of large scale projects, which can be accessed using standard machine communication protocols such as CQL, OAI-PMH *etc.* The function of Information Inference System Framework is to enrich the data stored in enterprise central data base, with the help of different data mining algorithms.

In this model the IIF contains the data flow nodes (Relevant Data, Contents Used, Content Analysis, Classification, Clustering, and Inferred Data), as shown in Figure 3. The data flow node named 'Relevant Data' retrieves the data from enterprise central data base. The second data flow node 'Content Used' finds out the similarities between the contents provided by the data source. The third node 'Content Analysis' compares the relationship between metadata of contents retrieved from enterprise central database. In

the 'Classification' node, the IIF assigns the labels to contents of large scale project, and then the 'Clustering' node generates the cluster of similar contents; the similarity depends on the metadata as well as functionality of that content in large scale project. Finally the node 'Inferred Data' exports the inferred data to enterprise central data base.

Together with the component nodes described and the information flows delineated, the proposed model is better poised to address the Big Data requirements in large scale complex project environments. We propose that the outlined model may be implemented to enhance communication productivity and efficiency in emerging networks and Big Data setups.

5. Conclusion

In this paper a new model is developed for communication management of large projects using IIF in Big Data Environment. Big Data are usually generated in various sectors of large scale projects, often distributed geographically across the country or world, and are characterized by huge volumes and varieties. Therefore, there is an immense need for platforms handling big amounts of data through complex event processing systems. Such systems should be able to handle multiple heterogeneous sources, deal with different issues regarding efficiently disseminating data, and collect and analyze them for enhanced performance of large scale projects. As Big Data is going to be part of every organization in near future, there is an evolving requirement to meet the communication needs of multiple stakeholders in large scale projects. The presented technological model can become the core of open source solution for processing huge data volumes within scattered heterogeneous projects anywhere in the world. The model connects different data sources within a large scale project to the enterprise data ware house and accesses the inferred data produced by the IIF framework. Compared to other structures, the IIF framework can process huge volumes of data in faster fashion. The features of this model can help different stakeholders involved in the large projects to efficiently access the data resources, best communicate for project execution, and exploit the data resources for successful completion of projects.

Acknowledgment

This work was supported under National Science & Technology Major Projects Process Monitoring Management Model and Empirical Research (National Soft Science Research Program Project Number 2013GXS3B060-02). We also acknowledge the kind collaboration offered by Prof. Dr. Wang ChangFeng at Beijing University of Posts and Telecommunications China.

References

- [1] S. R. Clegg, T. S. Pitsis, T. Rura-Polley and M. Marosszeky, "Govern mentality matters: Designing an alliance culture of inter-organizational collaboration for managing projects", *Organization Studies*, vol. 23, no. 3, (2002), pp. 317-337.
- [2] S. Rasheed, C. Wang and B. Lucena, "Risk Leveling in Program Environments—A Structured Approach for Program Risk Management", *Sustainability*, vol. 7, no. 5, (2015), pp. 5896-5919.
- [3] E. W. Merrow, "Understanding the outcomes of Mega projects: A quantitative analysis of very large civilian projects", *Rand*, (1988).
- [4] S. Moore and R. Shangraw Jr., "Managing Risk and Uncertainty in Large-Scale University Research Projects." *Research Management Review*, vol 18, no. 2, (2011), pp. 59-78.
- [5] S. Rasheed, C. Wang and F. Yaqub, "Towards Program Risk Management and Perceived Risk Management Barriers", *International Journal of Hybrid Information Technology*, vol. 8, no. 6, (2015), pp. 323-338.
- [6] R. R. Nelson, "IT project management: Infamous failure, classic mistakes, and best practices", *MIS Quarterly Executive*, vol. 6, no. 2, (2007), pp. 67-78.
- [7] A. Ahimbisibwe and S. Nangoli, "Project communication, individual commitment, social networks, and perceived project performance", *Journal of African Business*, vol. 13, no. 2, (2012), pp. 101-114.

- [8] V. Mayer-Schonberger and K. Cukier, "Big Data: A Revolution That Will Transform How We Live, Work, and Think", Houghton Mifflin Harcourt, (2013).
- [9] C. O'Neil and R. Schutt, "Doing Data Science: Straight Talk from the Frontline", O'Reilly Media, Inc., (2013).
- [10] D. Zeng and R. Lusch, "Big data analytics: perspective shifting from transactions to ecosystems", IEEE Intell. Syst., vol. 28, no. 2, (2013), pp. 2-5.
- [11] A. P. C. Chan, D. Scott and A. P. L. Chan, "Factors affecting the success of a construction project", J. Constr. Eng. Manag., vol. 130, no. 1, (2004), pp. 153-155.
- [12] D. M. Kennedy, S. A. McComb and R. R. Vozdolska, "An investigation of project complexity's influence on team communication using Monte Carlo simulation", J. Eng. Technol. Manag., vol. 28, (2011), pp. 109-127.
- [13] Y. Lu, L. Luo and H. Wang, "Measurement model of project complexity for large-scale projects from task and organization perspective", International Journal of Project Management, (2014).
- [14] K. M. Carley and W. A. Wallace, "Computational organization theory", Springer US, (2001).
- [15] W. Johnston and R. McCool, "The Square Kilometer Array—A next generation scientific instrument and its implications for networks", ESnet, Lawrence Berkeley National Laboratory, Signal Transport and Networks, SKA Program Development Office, Jodrell Bank Centre for Astrophysics, TERENA Networking Conference, (2012).
- [16] I. Monga, E. Pouyoul and C. Guok, "Software-Defined Networking for Big-Data Science-Architectural Models from Campus to the WAN", High Performance Computing, Networking, Storage and Analysis (SCC), SC Companion IEEE, (2012), pp. 1629-1635.
- [17] S. L. Brown and K. M. Eisenhardt, "Product development: Past research, present findings, and future directions", Academy of management review, vol. 20, no. 2, (1995), pp. 343-378.
- [18] A. Guide, "Project Management Body of Knowledge (PMBOK® GUIDE)", Project Management Institute, (2001).
- [19] Office of Government of Commerce, Managing Successful Programmes. Program Management Group, (2011).
- [20] M. Väänänen, "Communication in high technology product development projects", Project personnels viewpoint for improvement, Acta Universitatis Ouluensis C, vol. 364, (2010).
- [21] E. F. Harshman and C. L. Harshman, "Communicating with employees: Building on an ethical foundation", Journal of Business Ethics, vol. 19, no. 1, (1999), pp. 3-19.
- [22] M. Väänänen, "Communication in high technology product development projects", Project personnels viewpoint for improvement, Acta Universitatis Ouluensis C, vol. 364, (2010).
- [23] U. Alatalo, "Communication Strategy in Projects", High Technology Sector Viewpoint, (2012).
- [24] K. Bakshi, "Considerations for big data: Architecture and approach", Aerospace Conference, IEEE, (2012), pp. 1-7.
- [25] Index, Cisco Visual Networking, "Global mobile data traffic forecast update, 2010-2015", White Paper, (2011) February.
- [26] S. Lu, R. M. Li and W. C. Tjhi, "A framework for cloud-based large-scale data analytics and visualization: Case study on multiscale climate data", Cloud Computing Technology and Science (CloudCom), IEEE Third International Conference on IEEE, (2011), pp. 618-622.
- [27] I. Mierswa, M. Wurst and R. Klinkenberg, "Yale: Rapid prototyping for complex data mining tasks", Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, (2006), pp. 935-940.
- [28] P. Manghi, M. Artini, C. Atzori, A. Bardi, A. Mannonci, S. La Bruzzo, L. Candela, D. Castelli and P. Pagano, "The D-NET software toolkit: A framework for the realization, maintenance, and operation of aggregative infrastructures", Journal Program, Emerald Insight, vol. 48, no. 4, (2014), pp. 322-354.

Authors



Suhail Memon, is a Ph.D. researcher at Beijing University of Posts and Telecommunications, China. He has been affiliated to the education department since 2008. Before commencing his doctoral research, he worked as Teacher at Government High School in Pakistan. In 2005 He received BCS (Hons) From Quaid-e-Awam University of Engineering Science & Technology Nawabshah Sindh Pakistan. He received MBA(MIS) from University of Sindh Pakistan in 2008. His areas of research are Communication Management, Project Management, and Managing Information Systems.



Wang ChangFeng, is the director of the International Project Management Institute in Beijing University of Posts and Telecommunications, professor, and Ph.D. supervisor. He is an expert in guiding the PMI GAC project management accreditation, assessment expert for PMI GAC CRC project management accreditation. His main research areas include the enterprise project management, safety risk early warning and emergency response, the complex system integration, and control of major projects and programs.



Shahid Rasheed, is a Ph.D. researcher at Beijing University of Posts and Telecommunications, China. He has been affiliated to the telecommunications, defence and other engineering sectors since 1998. Before commencing his doctoral research, he worked as Senior Manager at Pakistan Telecommunication Company Ltd. (Etisalat), mainly in the projects and strategic management functions. He holds post-graduation qualifications in Business Administration as well as Telecom Engineering fields. His areas of interest include Program and Strategy Management, ICT developments, and allied disciplines.



Zulfiqar Hussain Pathan, is a Ph.D. scholar at Beijing University of Posts and Telecommunications, China. He has been affiliated to Mehran University of Engineering & Technology, (MUET) Jamshoro, Pakistan. He got his post-graduation in the field of Technology & Innovation Management and graduation in Computer Science. His research interests are Information System, Innovative & Technological practices in Management in Public & Private Sectors, and Entrepreneurship.



Qiu Yixin, is a Master researcher at Beijing University of Posts and Telecommunications, China. She has been affiliated to the Education since 2014. Before commencing her master's research, she studied in Jiangxi University of Science and Technology during 2010 to 2014. In 2014 she received bachelor's degree from Jiangxi University of Science and Technology. During her bachelor education, she won National Scholarship, Top Ten Graduate, Provincial "triple-A" outstanding student, the second award of national mathematics competition and some other prize.



Liu Yanping, is a Master researcher at Beijing University of Posts and Telecommunications, She majored in Management Science and Engineering, Her current area of research is Project Management. She is the president of project management innovation institution Beijing University of Posts and Telecommunications.