

Data Dissemination for Bioinformatics Application using Agent Migration

Shakir Ullah Shah¹

Iqra University, Islamabad, Pakistan
National University of Computer and Emerging Sciences
Peshawar, Pakistan

Abdul Hameed²

Department of Computer Science
Iqra University, Islamabad, Pakistan

Jamil Ahmad³

Senior IEEE Member Hazara University
Mansehra KP
Pakistan

Hafeez Ur Rehman Safia Fatima⁴, Muhammad Amin⁵

National University of Computer and Emerging Sciences
Peshawar, Pakistan

Abstract—Bioinformatics is research intensive field where agents operate in highly dynamic environment. Due to extensive research in this domain leads to basic but important problems for the researchers that are (1) Bandwidth (2) storage and (3) computation. We are using agent migration approach to reduce the network load and resolve the resource problem for the client by using server side resources for the computations on large data. The proposed approach does not demand extra storage and extensive computational resources on clients side fsage. It solves the problem of bandwidth, storage, computation. Our results show that this approach saves the time of the user up to 12.5 % approximately, depending on the size of the data. Similarly the agent can work like a mashup to get heterogeneous data from different service providers and presents in homogeneous shape to its owner.

Keywords—Data dissemination; protein-protein interactions; agent migration; inter-platform mobility; multi-agent systems

I. INTRODUCTION

Bioinformatics [1], [2] is one of the applications of computer science for managing the biological information. It is interdisciplinary field of sciences which combines mathematics, computer science, statistics, and engineering to understand bio- logical data. It is a vast field and with the number of researcher doing research in it makes it more important. Researchers have done a lot of work to find and understand the nature as well as dynamics of proteins [3]. The protein-protein interactions [4], [5] involve structural information [4] of proteins as well as the non-structural information [6], [7]. A variety of computational approaches [8] have also been developed. Yet comparative studies suggest that the data involve in it is large, complex and not in one standard format; with Next Generation Sequences (NGS) [9], [10], [11] on which researcher are doing work to extract new interactions, using this (NGS) leads to another problem that is storage [12], [13].

There are a lot of experimental methods [14] which are noisy, costly in term of computation and storage and time consuming to predict the protein-protein interactions. Due to high computation, high storage need and data heterogeneity it is hard for researchers to carry out their research, so here we

recommend an agent [15], [16], [17] based approach which will reduce the bandwidth need, transfer computations to the machine who has high computational power and will give data in homogeneous format to increase the researchers' productivity. Although agent itself has many characteristics but the characteristics we will be using throughout this study is agent mobility. The big picture is that an agent gets requested from its owner and visits service provide as per the list to fetch required data, manipulate the data at service side and return back to its originated platform with the results. In this study, we are not targeting on communication problems and assume that environment is up and running.

The rest of the paper is organized as follows. In section II, literature about the domain is given. Proposed solution along with main steps is given Section III. Detail about reference implementation is given in Section IV for the proof of concept. The importance of the proposed solution is given in the form of results are discussed in Section V. The implication of the research work is in Section VI.

II. LITERATURE REVIEW

Proteins are large molecules, which are the collection of amino acids, are essential to our bodies to function properly [18]. Proteins are very important component in the proper functioning and maintaining of our body structures, its normal functions and the regulation of the body's different parts. Enzymes which are responsible to speed up a chemical reaction are also proteins. Oxygen is an essential element for all living being for their survival and proteins play utmost important role as a carrier in the form of hemoglobin. Proteins help us fight infection as well as DNA the building blocks to life. It's too required to create up muscle tissue, which in turn makes a difference to keep our bodies dynamic, solid, and healthy. Most protein is put away within the body as muscle, by and large bookkeeping for around 40-45% of our bodies add up to pool [6].

Researchers have done a huge amount of work to find and understand the nature as well as changing aspects of proteins. The protein-protein interactions [4], [5], [19] involve either structural information of proteins like Domains, 3-D shape of

proteins, structural neighbors as well as the non-structural information that includes protein homology, sequence similarity, functional similarity etc. Extended form of computational approaches, for illustration, on arrangement homology, quality co-expression and phylogenetic profiles, have moreover been created for the genome-wide deduction of protein-protein interactions (PPIs).

Prediction of PPIs at the structural level is essential as it allows predication of protein functions, helps in the discovery of drug and so play vital role in so many other areas [6], [20]. Protein Interactions by Structural Matching (PRISM) [21], [22] protocol is provided huge scale forecast of protein-protein intuitive and gathering of protein complex structure. PRISM method consists of two parts:

- Firm body basic structural comparison of the intended protein to know the template PPIs.
- Adaptable refinement through the use of docking energy function.

PRISM predicts binding residue by using structural likeness and developmental conversation of putative binding residue but require high computational power and high bandwidth to stay active. Huge number of tools and models have been developed in recent years for the interpretation of biological data, but not all of these are publicly available or permit bulk submission via web [23]. While few tools and models require proper training and background knowledge but the proposed solution is very simple.

There is a huge growth in the biological sequence where a tremendous sum of information is being created and uploaded on the web sites/servers. Now to get the data we would need to interact with the interface using web based queries [24]. This means that the researcher has to do a query each time he/she needs the data source. Above all these resources would be in different formats, entries, query options etc. Moreover, this process requires the researchers to remain online and wait for the required results. Secondly this approach needs high bandwidth. This is also very important retrieve the results on low computation powered resources like mobile phones. This study propose migration feature of the agent to overcome aforementioned problems by transforming the computation at the required host of resources [17].

Knowledge administration is a repetitive, complex and time-consuming task. It requires high computational resources. In specific, the kinds of assets accessible within the bioinformatics space are various databases and investigation instruments. These resources can be autonomously managed in topographically unmistakable areas, using Multi-agent approach [25]. Researchers, in the field of bioinformatics, consistently propose various techniques to resolve such issues.

There are various approaches to retrieve data from server like remote procedure call (RPC), java Remote Method Invocation (RMI), etc. [26], [27]. This study focus only on using Multi-Agent Systems (MAS) [16]. MAS, like Jason [28] is a system comprised of multiple agents, a peace of computational logical unit to perform different tasks on behalf of its creator.

III. PROPOSED SOLUTION

We propose Agent migration approach to solve the problems of low Bandwidth, low storage extensive resources and dynamic environment. The purpose of using agent migration approach is to minimize the network load, increase flexibility and enhance parallel processing. Mobile agents are actually autonomous (act independently) programs that do travel form one system to another in a network. Mobile agents are proactive, reactive, flexible and social. They are trained to the task assigned the users. Due to their functionalities as mentioned above , they proved highly effective many scenarios where everyone is busy and have less time and more task to do[10, 11]. They have the capability that they suspend their execution in one system and migrate to other system to resume their computation. To consume less time suspension strategy has proven its worth, by suspending when one system migrating to another system to resume the work is highly effective when previous system doesn't have required sources to complete the task, agent move to another system with the task, complete the task over their on second system and get to fist system with results only. This way all the users don't have to buy a high specifications system, they can use other high power system to complete their work.

A. Agent Mobility

In agents migration, agent mobility has further two types: Inter-platform mobility and Intra-platform mobility. In case of intra-platform mobility agent moves between different containers within the same platform on the other hand in inter-platform mobility agent leave one system lets say client and move or migrate to other system so called server, means agent is moving between different platforms. The main focus of this is intra-platform mobility.

B. Main Steps of Proposed Solution

This study proposes a step by step solution based on agent mobility. The pre and post details of each step provide great insight of the requirements. These steps are visualized in Fig. 1 and its sequence diagram is given in Fig. 2.

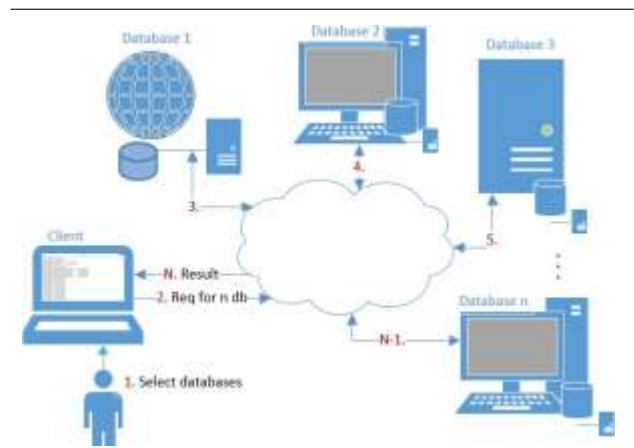


Fig. 1. Main Steps of Proposed Solution.

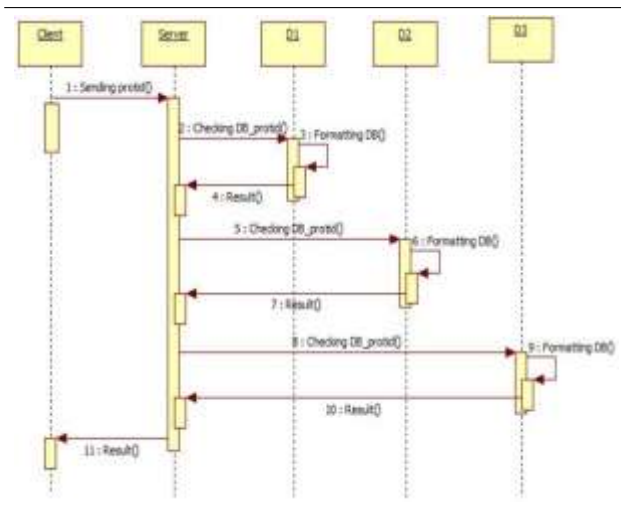


Fig. 2. Sequence Diagram.

The detail of each step is given below:

- 1) *Enter protein ID*: Here the user will enter the protein id of which he/she wants to check the protein to protein Interactions.
- 2) *Select dbs*: On the bases of given id, all possible protein to protein interactions will be shown to user. User will select databases according to its need.
- 3) *Fetching the data*: Agent will leave the client and will fetch the data from each service provider.
- 4) *Merge all in one file*: As the user can select multiple databases where the interactions would be in different formates. Agent will convert heterogeneous formates into homoge- neous formate, like xml.

IV. REFERENCE IMPLEMENTATION

There are various agent framework to deploy agents [29],[30] In this study JADE [31] is used as an agent framework as it is fully complaint to Foundation for Intelligent Physical Agents (FIPA) [32], open source and is used in the state of art agent technology. JADE does not support intra-platform mobility, so for that JIPMS [33] is used for this purpose. It is also based on JAVA, so it becomes fare to compare it with JAVA RMI. For the sack of comparison, protein id and selection of other parameter like number of interactions remained same for both. The main GUI can be seen in Fig. 3. When the user enters protein id and number, then the agent fetches the record of possible protein to protein interaction from all available databases and gives a table view. The user will select all those dbs that are required for her. When the user complete its selection, then the agent will fetch all the data from selected databases, Agent will combine all required data in a predefined format. All the files or interactions would be downloaded in XML format with different tags i.e. Source db, author etc. Then agent comes back to its originator platform. When agent notifies the user about the work done, then the user can view the fetched data on client side.

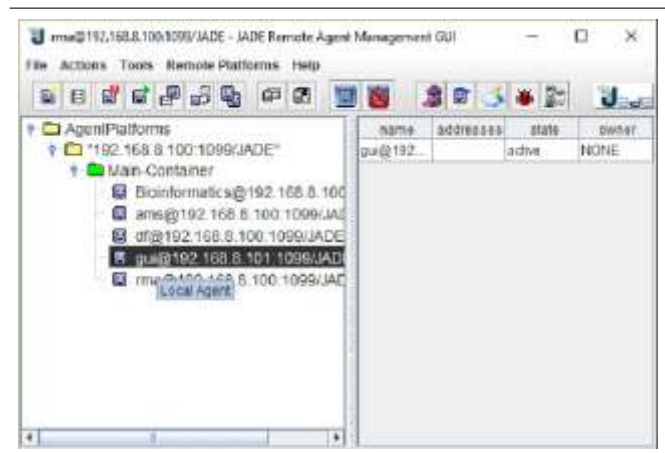


Fig. 3. Main GUI.

V. RESULT AND DISCUSSION

This section describes the results achieved so far and concludes with some discussion. There are different studies carried to compare performance between JAVA RMI and agent based system against response time and network load [?]. The study focused on implementing proposed method using JAVA RMI and mobile agent framework. The Fig. 4 shows that agent based approach causes less load over the network as compared to JAVA RMI.

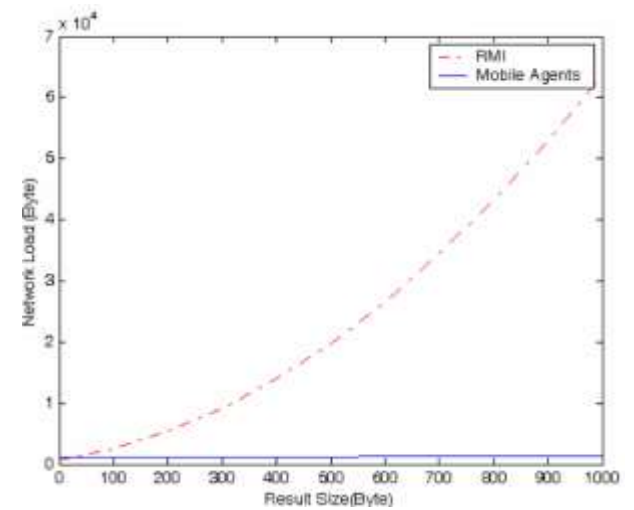


Fig. 4. Client Side Network Load and Result Size.

Fig. 5 shows relation between response time and result size. It shows that agent based approach produce better results as compared to JAVA RMI when the result size becomes greater than 2KB.

In the Fig. 6 given above the blue line shows agents graph and red line shows the graph of java RMI we can clearly see from the graph that if we decrease the bandwidth our agent is computing faster as compare to JAVA RMI. As agent can move at low bandwidth and don't require high bandwidth for migration and so on that's why agent is showing good results even at low bandwidth unlike JAVA RMI. Moreover JAVA RMI won't even work when bandwidth is 5Mbps unlike agent;

agent can migrate at 5 Mbs. If let's say the bandwidth is 15Mbs then our agent takes almost 41.66% (time/60*100) to serve the clients request while for the same bandwidth java is taking more time than agent. Using agents has solved the bandwidth problem, Storage problem and the formatting too.

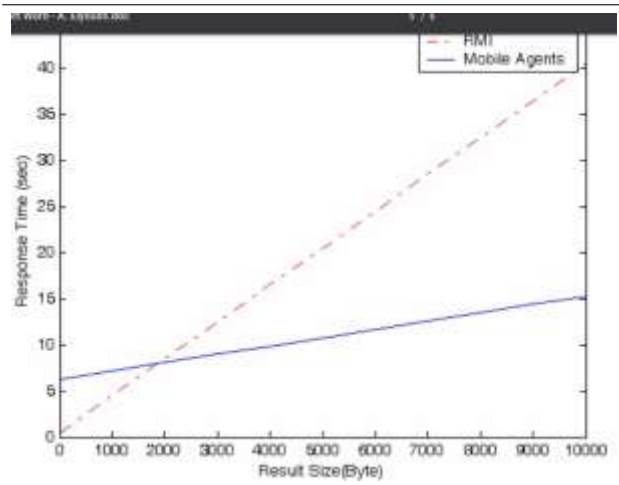


Fig. 5. Response Time vs Result Size.

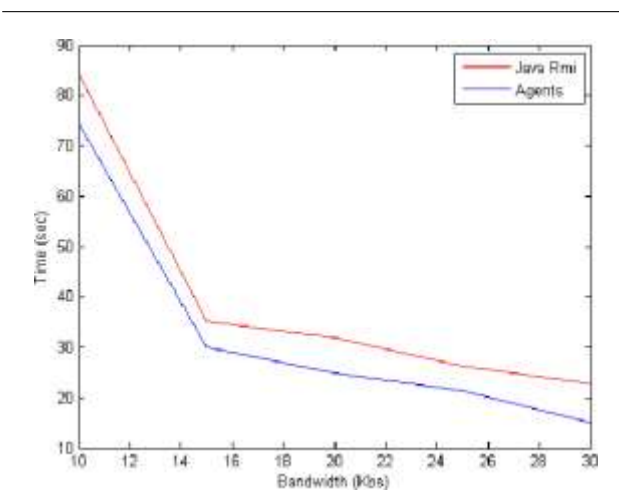


Fig. 6. Comparison between Agents and Java RMI.

VI. CONCLUSION AND FUTURE WORK

Mobile Agent technology can be used in many areas, merging this technique with other areas can also be useful because it does not require high bandwidth or strong Internet connection. Agents are intelligent and can even work well even in low network areas. If we see the future of agents' migration approach in bioinformatics it can be highly beneficial for huge amount of data and more computations. It can be used for many generic purposes as well. This study found the interactions between proteins using agent migration protocol. This approach found that client with limited resources can also be used for finding protein to protein interaction. The finding of this study is that mobile agent technology leverage network load and storage on client side and heterogeneous data can be converted into homogeneous format. Furthermore this approach does not demand the

availability of the user online for full time. Our research can be modified to make it work on different bioinformatics problem like viewing the interaction of sequences.

ACKNOWLEDGMENT

We are grateful to Waqas Haider Khan Bangyal for reading the manuscript and for improving proofs of this manuscript several times.

REFERENCES

- [1] A. D. Baxevanis, G. D. Bader, and D. S. Wishart, *Bioinformatics*. John Wiley & Sons, 2020.
- [2] R. Stevens, C. A. Goble, and S. Bechhofer, "Ontology-based knowledge representation for bioinformatics," *Briefings in bioinformatics*, vol. 1, no. 4, pp. 398–414, 2000.
- [3] A. Amadei, A. B. Linssen, and H. J. Berendsen, "Essential dynamics of proteins," *Proteins: Structure, Function, and Bioinformatics*, vol. 17, no. 4, pp. 412–425, 1993.
- [4] D. F. Waugh, "Protein-protein interactions," *Advances in protein chemistry*, vol. 9, pp. 325–437, 1954.
- [5] T. Bergg a'rd, S. Linse, and P. James, "Methods for the detection and analysis of protein-protein interactions," *Proteomics*, vol. 7, no. 16, pp. 2833–2842, 2007.
- [6] Q. C. Zhang, D. Petrey, L. Deng, L. Qiang, Y. Shi, C. A. Thu, B. Bisikirska, C. Lefebvre, D. Accili, T. Hunter et al., "Structure-based prediction of protein-protein interactions on a genome-wide scale," *Nature*, vol. 490, no. 7421, p. 556, 2012.
- [7] S. Das and S. Chakrabarti, "Classification and prediction of protein-protein interaction interface using machine learning algorithm," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021.
- [8] K. A. Theofilatos, C. M. Dimitrakopoulos, A. K. Tsakalidis, S. D. Likothanassis, S. T. Papadimitriou, and S. P. Mavroudi, "Com- putational approaches for the prediction of protein-protein interactions: a survey," *Current Bioinformatics*, vol. 6, no. 4, pp. 398–414, 2011.
- [9] R. B. Russell, F. Alber, P. Aloy, F. P. Davis, D. Korkin, M. Pichaud, M. Topf, and A. Sali, "A structural perspective on protein-protein interactions," *Current opinion in structural biology*, vol. 14, no. 3, pp. 313–324, 2004.
- [10] B. Suter, X. Zhang, C. G. Pesce, A. R. Mendelsohn, S. P. Dinesh- Kumar, and J.-H. Mao, "Next-generation sequencing for binary protein-protein interactions," *Frontiers in genetics*, vol. 6, p. 346, 2015.
- [11] R. Carter, A. Luchini, L. Liotta, and A. Haymond, "Next-generation techniques for determination of protein-protein interactions: Beyond the crystal structure," *Current pathobiology reports*, vol. 7, no. 3, pp. 61–71, 2019.
- [12] D. S. Horner, G. Pavesi, T. Castrignano, P. D. De Meo, S. Liuni, M. Sammeth, E. Picardi, and G. Pesole, "Bioinformatics approaches for genomics and post genomics applications of next-generation sequenc- ing," *Briefings in bioinformatics*, vol. 11, no. 2, pp. 181–197, 2010.
- [13] J. K. Kulski, "Next-generation sequencing overview of the history, tools, and omic applications," *Next generation sequencing-advances, applications and challenges*, pp. 3–60, 2016.
- [14] N. Safari-Alighiarloo, M. Taghizadeh, M. Rezaei-Tavirani, B. Goliaei, and A. A. Peyvandi, "Protein-protein interaction networks (ppi) and complex diseases," *Gastroenterology and Hepatology from bed to bench*, vol. 7, no. 1, p. 17, 2014.
- [15] M. J. Wooldridge and N. R. Jennings, "Intelligent agents: Theory and practice," *The knowledge engineering review*, vol. 10, no. 2, pp. 115–152, 1995.
- [16] A. Omicini, A. Ricci, and M. Viroli, "Artifacts in the a&a meta-model for multi-agent systems," *Autonomous agents and multi-agent systems*, vol. 17, no. 3, pp. 432–456, 2008.
- [17] L. Gao, H. Dai, T.-L. Zhang, and K.-C. Chou, "Remote data retrieval for bioinformatics applications: An agent migration approach," *PLoS one*, vol. 6, no. 6, p. e20949, 2011.
- [18] T. Simonson, "Electrostatics and dynamics of proteins," *Reports on Progress in Physics*, vol. 66, no. 5, p. 737, 2003.

- [19] M. Shatsky, R. Nussinov, and H. J. Wolfson, "A method for simultaneous alignment of multiple protein structures," *Proteins: Structure, Function, and Bioinformatics*, vol. 56, no. 1, pp. 143–156, 2004.
- [20] J. Zahiri, J. Hannon Bozorgmehr, and A. Masoudi-Nejad, "Computational prediction of protein–protein interaction networks: algorithms and resources," *Current genomics*, vol. 14, no. 6, pp. 397–414, 2013.
- [21] U. Ogmen, O. Keskin, A. S. Aytuna, R. Nussinov, and A. Gursoy, "Prism: protein interactions by structural matching," *Nucleic acids research*, vol. 33, no. suppl 2, pp. W331–W336, 2005.
- [22] N. Tuncbag, A. Gursoy, R. Nussinov, and O. Keskin, "Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using prism," *Nature protocols*, vol. 6, no. 9, p. 1341, 2011.
- [23] Y. Ding and L. Gao, "Macrodynamics analysis of migration behaviors in large-scale mobile agent systems for the future internet," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 5, pp. 1032–1036, 2011.
- [24] K.-C. Chou, "Some remarks on protein attribute prediction and pseudo amino acid composition," *Journal of theoretical biology*, vol. 273, no. 1, pp. 236–247, 2011.
- [25] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.
- [26] G. A. Aderounmu, B. Oyatokun, and M. Adigun, "Remote method invocation and mobil agent: A comparative analysis." *Issues in Informing Science & Information Technology*, vol. 3, 2006.
- [27] K. Miller and G. Mansingh, "Comparing the use of mobile intelligent agents vs client server approach in a distributed mobile health application." *JCP*, vol. 10, no. 6, pp. 365–373, 2015.
- [28] R. H. Bordini and J. F. H u' bner, "Bdi agent programming in agentspeak using jason," in *International workshop on computational logic in multi-agent systems*. Springer, 2005, pp. 143–164.
- [29] R. H. Bordini, L. Braubach, M. Dastani, A. E. F. Seghrouchni, J. J. Gomez-Sanz, J. Leite, G. O'Hare, A. Pokahr, and A. Ricci, "A survey of programming languages and platforms for multi-agent systems," *Informatica*, vol. 30, no. 1, 2006.
- [30] K. Kravari and N. Bassiliades, "A survey of agent platforms," *Journal of Artificial Societies and Social Simulation*, vol. 18, no. 1, p. 11, 2015.
- [31] F. Bellifemine, "Jade-a white paper," *exp*, vol. 3, no. 3, 2003.
- [32] T. SpA. (2001) The foundation for intelligent physical agents. <https://www.fipa.org/>. [Online]. Available: <https://www.fipa.org/>.
- [33] C. J. (2006) Inter-platform mobility project. Senda research group of Autonomous University of Barcelona. [Online]. Available: <https://tao.uab.cat/ipmp/>.