International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

Energy efficient Approach of big data gathering in wireless sensor networks

[1] Kummarikuntla Lakshmi Mohan Kumar; [2] B S. Madhuri & [3]S.Neelima

[1]M.Tech (VLSI & Embedded System),
[2]Asst.Professor,
[3]HOD, Assoc.Proffesor,
Gandhiji Institute of Science & Technology (Gist) – JNTU Kakinada Affiliated,
Gattubhimavaram (Village),Jaggayyapet (Mandal),Krishna (District),Andhra Pradesh, India.

Abstract-

Recently, the "big data" emerged as a hot topic because of the tremendous growth of the Information and Communication Technology (ICT). One of the highly anticipated key contributors of the big data in the future networks is the distributed Wireless Sensor Networks (WSNs). Although the data generated by an individual sensor may not appear to be significant, the overall data generated across numerous sensors in the densely distributed WSNs can produce a significant portion of the big data. Energy-efficient big data gathering in the densely distributed sensor networks is, therefore, a challenging research area. One of the most effective solutions to address this challenge is to utilize the sink node's mobility to facilitate the data gathering. While this technique can reduce energy consumption of the sensor nodes, the use of mobile sink presents additional challenges such as determining the sink node's trajectory and cluster formation prior to data collection. In this paper, we propose a new mobile sink routing and data gathering method through network clustering based on modified Expectation- Maximization (EM) technique. In addition, we derive an optimal number of clusters to minimize the energy consumption.

Index Terms- Big data; Wireless Sensor Networks (WSNs); clustering; optimization; data gathering; and energy efficiency

1 INTRODUCTION

Big data, as a concept, was first proposed by META Group analyst Doug Laney in the 2001 research report [1] and his related lectures. Increasing volume (amount of data), velocity (speed of data), and variety (range of data types and sources) are used as three important characteristics to define big data. As for now, two new characters, value and veracity, are added by some organizations [2] to further illustrate the necessary properties of big data. This "5Vs" model, which is used for describing big data and its related challenges, like data capture, storage, search, sharing, transfer, analysis, and visualization, is a hot topic in current data science research field. In the field of sensing, special issues are generated. With the exponential increasing number of data generating devices (such as computers, tablets, and sensors, especially smart phones), vast amount of data needs to be processed. Research methods for big data can be applied to various fields by utilizing sensing techniques, such as science, engineering medicine, health care, finance, business, and ultimately the whole society. However, currently, there is still no generic and systematic big data research model in the world of sensing. The vision of data processing in future sensing is vague and relevant infrastructures and structures have not yet been well defined.

The most intuitive understanding that comes into people's mind is a large amount of data reflecting the space domain of data sourcing. In the 5Vs model, volume and variety are directly relevant to this understanding. In the world of sensing, large amount of data is usually gotten from a large sensing area, for example, town or city level sensing or the applications for Internet of Things. Town or city level sensing relies not only on sensors within city infrastructures, but also on a large number of device owners willing to sense and contribute their data to data aggregation platforms. A survey result shows that every day we create more than 2.5 quintillion bytes of data, and a prediction says that, in 2016, over 4.1 terabytes of data will be generated per day per square kilometer in urbanized land area. Furthermore, in 2016, it is estimated that 39.5 billion dollars will be spent on smart city technologies, up from 8.1 billion dollars in 2010 [3]. The pervasive use of mobile phones and other similar mobile sensing devices will account for a dominant portion of aforementioned Increment. Smart phones enable everyone to collect data at any time and place. Although some sensing data may not be valuable to the sensor owner, they can be valuable to the scientific community. Currently, building a generic sensing platform for a city scale data application faces many challenges. The first challenge is how to design a system in which users can benefit from data sharing . As one of the most important parts of city scale sensor, personal sensing devices are still within the "owner-is-the-user" model. Getting considerable benefits without personal information leakage is the baseline of making full use of individual sensing data, as privacy and security are general concerns. The second challenge is how to effectively collect the data scattered in the individual sensing devices. The large amount of data generated by distributed sensors typically does not have a central control or a centralized accounting device that can be notified when new data is generated.

The variety indicates that the data is of highly varied structures (e.g. data generated by a wide range of sources such as Machine-to-Machine (M2M), Radio Frequency Identification (RFID)[4], and sensors) while the velocity refers to the high speed processing/analysis (e.g., click-streaming, fast database transactions, and so forth). On the other hand, the volume refers to the fact that a lot of data needs to be gathered for processing and analysis. Gathering the large volume and wide variety of the sensed data is, indeed, critical as a number of important domains of human endeavor are becoming increasingly reliant on these remotely sensed information. For example, in smart-houses with densely deployed sensors, users can access temperature, humidity, health information, electricity consumption, and so forth by using smart sensing devices.

In order to gather these data, the Wireless Sensor Networks (WSNs) are constructed whereby the sensors relay their data to the "sink". However, in case of widely and densely distributed WSNs (e.g. in schools, urban areas, mountains, and so forth) [5], there are two problems in gathering the data sensed by millions of sensors. First, the network is divided to some sub-networks because of the limited wireless communication range. For example, sensors deployed in a building may not be able to communicate with the sensors which are distributed in the neighboring buildings. Even though the volume of data generated by an individual sensor is not significant, each sensor requires a lot of energy to relay the data generated by surrounding sensors. Especially in dense WSNs, the life time of sensors will be very short because each sensor node relays a lot of data generated by tremendous number of surrounding sensors. In order to solve these problems, we need an energy-efficient method to gather huge volume of data from a large number of sensors in the densely distributed WSNs.

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

2. Overview of IEEE 802.15.4

2.1. Device Types

Two different types of devices are defined in an IEEE 802.15.4 network. A Full Function Device (FFD) supports the protocols of the wireless standard defined for WSN. A Reduced Function Device (RFD) provides limited functionality with the low cost and complexity. It is generally used for the network edge devices with very low power consumption. When using a star topology, a unique central FFD device is defined and operates as a PAN coordinator. This coordinator will manage the whole network and all data transmitted to any destination have to pass through it. In consequence, it required much more memory and power.

2.2. IEEE 802.15.4 and ZigBee Relationship

IEEE 802.15.4 is part of the IEEE family of standards for physical and MAC layers for Wireless Personal Area Networks (WPANs). The focus area of IEEE 802.15.4 is that of low data rate WPANs, with low complexity and stringent power consumption requirements. The standard is designed to be cost effective, low-power, and low-interference technology. This technology operates on the same 2.4GHz ISM band as WiFi, Bluetooth, and WiMax. Depending on the environment and the power used for transmission, IEEE 802.15.4 compliant wireless devices are expected to transmit in a range of 10 to 75 meters. The ZigBee Alliance is a group of over 400 member companies that maintain and publish technical standards. ZigBee is a registered trademark of ZigBee Alliance. IEEE defines the physical and medium access layers while the ZigBee [6] Alliance defines network and application layers. The group recognizes that the coexistence of different wireless technologies on the same frequencies can have a significant impact on network operations.

2.3. The MAC Layer

The 802.15.4 standard defines data link layer (DLL) and physical layer (PHY) protocols for supporting sensor devices with minimal power consuming and operating in limited area. The two layers provide the reliable communication between the nodes of the network by avoiding the collisions to improve the efficiency [6]. The DLL is divided into two sub layers, the medium access control (MAC) and logical link control (LLC) sub layers. The LLC is standardized in IEEE 802.2 and is common among the IEEE 802 standards. The MAC layer provides an interface between upper layers and the PHY layer. It is responsible for frame validation using acknowledged frame delivery. It is also in charge of maintaining nodes synchronization, controlling the association, and choosing the Guaranteed Time Slot mechanism. Moreover, the MAC layer handles channel access employing the CSMA/CA [7] mechanism. The CSMA/CA protocol is an important mechanism for channel access, considering low data rate adopted in 802.15.4. This mechanism evaluates the channel and allows data packets to be transmitted if the condition is suitable (free of activities). Otherwise the algorithm shall back off for certain periods before assessing the channel again. Details of MAC layer characteristics are presented in the following paragraphs.

3 CLUSTERING-BASED BIG DATA GATHERING IN HEAVILY DISTRIBUTED WSN

3.1 Clustering problem

When considering the scheme of data gathering in WSN using mobile sink, the biggest challenge in reducing energy consumption is how to decide the location where data gathering is conducted. As we assume that required energy for data transmission of node is proportional to the square of transmission distance, the best clustering

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

algorithm to minimize energy consumption for data transmission must minimize the sum of square of data transmission distance in a network. EM algorithm is powerful and well-known tool to solve the clustering problem by repeatedly calculate the simple math formula. Since the EM algorithm can minimize the sum of square of distance between every node and cluster centroid. we adopt EM[8] algorithm over the 2-dimensional Gaussian mixture distribution. However, there is a limitation of the maximum communication range in the realistic situation. Not all nodes can connect to each other and also to the cluster centroid. Nodes that cannot directly communicate with the cluster centroid need to communicate in a multi-hop manner. In multi-hop communication, communication distance is a sum of distance between nodes in multi-hop path.

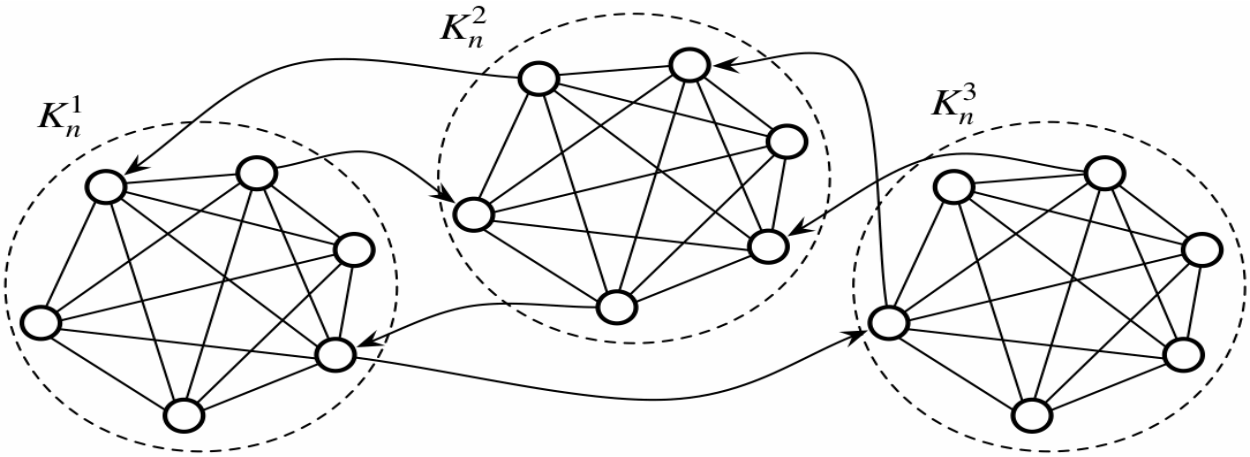

Fig 1 Clustered network.

3.2 Data gathering procedure using the proposed clustering technique

After clustering, the mobile sink patrols every cluster centroid and collects the data from the nodes in the cluster. It is easy to see that delay is a main problem of using mobile sink in WSNs. This delay is the waiting time from data generation to data sending. Because the mobile sink moves relatively slow compared with electrical communication between nodes, the mobile sink scheme causes long delay. To shorten this delay, we need to minimize total patrolling path length. Thus, in our scheme the mobile sink patrols along Traveling Salesman Problem (TSP) path of all cluster centroids. Once the mobile sink arrives at the cluster centroids, it collects data from sensor nodes. Directed Diffusion [9] is one of the most famous data collection schemes in WSNs. In our method, we consider using a typical example of them, "One Phase Pull [10] where the mobile sink node sends data request message at the cluster centroids. When a sensor node receives a data request message from cluster k, the node re-broadcasts the data request message and replies data to the neighboring node, which is the parent node in the data request tree of cluster k. Then, the node relays data messages to the sink.

To minimize the total required energy to send data, all nodes send the sensed information according to the value of responsibility of the cluster. The responsibility value is calculated based on the given parameters, and according to (3). These parameters are added to data request message and sent by the sink. Only after the sensor deployment, each node exchanges its own position vector, x, with sensor nodes belonging to same groups. Because the exchange of position vector is executed only one time after the sensor deployment, the energy consumption is not significant. As a result, when a node belongs to only one cluster, the node can send all data to the sink node. And when a node belongs to more than one

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

clusters, the node sends data according to the responsibility of each cluster.

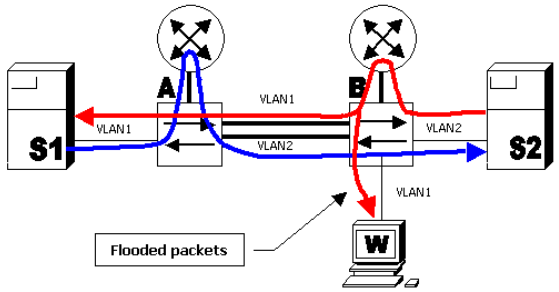

Fig. 2 Data request flooding in low and high connectivity network.

4 THE PROPOSED STRATEGY

Planetary is a light-weight, database oriented data aggregation system which is platform independent and focuses on energy efficiency. We want to focus on the data aggregation strategies and therefore the used routing strategy is predefined and arbitrary. Planetary does not enforce a single routing strategy but can support in finding optimal routes through the sensor network. In general, the database-oriented approach consists of two phases, similar to those in relational database systems Query compilation / optimization, Query execution / runtime. In our approach we call these two phases query propagation and aggregation, since the query has to be delivered to all concerned network nodes after its statement. After that, information at the nodes has to be aggregated and fused in network and the results need to be sent back to the client. It should be noted that query propagation and aggregation may be executed in parallel. If a query reaches a leaf node the results are immediately sent back if the query conditions are met. Also the propagation (as well as aggregation) itself is parallel since the query (or its results) are transferred into the sensor network similar to a breadth first search.

A WSN is composed of sensor nodes which are placed to monitor physical phenomena. Each node defines a wireless device having sensing, communication, and computation capabilities. Since WSN nodes are powered by non rechargeable battery, network life time is limited. In order to improve energy saving and to enhance lifetime of the network, it is necessary to ameliorate functions such as communication protocol, access channel mechanism, and routing. In particular, the slotted CSMA/CA mechanism in beacon enabled mode is improved to overcome a number of issues which may cause serious performance degradation. Degradation factors are related to network and traffic parameters, as might be expected, and in some cases these parameters pose severe limitations on the throughput and packet delays in such networks. The goal of this work is to analyze those issues and their impact and to suggest modifications of the CSMA/CA algorithm that offer much improved traffic performances.

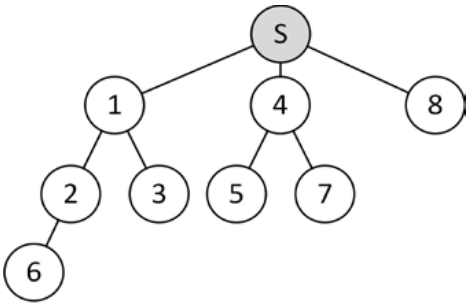

Fig.3 Network topology used for the evaluation

We present an original work that is based on implementing the modified MAC protocol on VHDL environment. Our work is focused particularly on studying the adaptive back off exponent (BE) management of CSMA/CA for 802.15.4. The reduced value of BEs may lead to the choice of identical BEs for different nodes. In consequence these nodes may wait for the same number of back off slots. This often leads to degradation of system performance at congestion scenarios, due to higher number of collisions. This paper addresses the problem by proposing an adaptive mechanism to the current implementation of the back off exponent management, based on a queue length criterion. Hence, when a queue size of a node increases over a fixed queue level, the parameter BE will be decreased to obtain a shorter back off period.

4.1 CSMA/CA

The Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) is widely employed in

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

wireless networking due to its simplicity and performance efficiency. It has been adopted as a medium access control protocol by many standards such as the IEEE 802.11 Wireless Local Area Networks (WLANs), IEEE 802.15.3, and IEEE 802.15.4 Wireless Personal Area Networks (WPANs). The slotted CSMA/CA shall first initialize the NB, the CW, and the BE. The BE shall be initialized to the value of macMinBE. Then, the slotted CSMA/CA locates the boundary of the next back off period. It shall delay for a random number of complete back off periods. Next, it requests that the PHY performs a CCA. This last shall start on a back off period boundary. In detail, after back off counter expires (reaches zero), the node must perform two clear channel assessments (CCAs) before trying to transmit. The two CCA operations ensure prevention of potential collisions. Transmission occurs if both CCAs are successful (sense the channel idle). If the channel is assessed to be busy, the MAC sub layer shall increment both the NB and the BE by 1, ensuring that BE shall be no more than macMaxBE. Furthermore, in CSMA/CA mode IEEE 802.15.4 supports optional retransmission scheme based on acknowledgements. When retransmissions are enabled, the receiver node must send a positive acknowledgement right after receiving a data frame. If the acknowledgment is not (correctly) received by the sender, a retransmission is started unless the maximum number of retransmissions is reached. In this case, the data frame is dropped. The maximum retransmission retries can be set between 0 and 7 with a default value of 3. If a collision occurs, CSMA/CA algorithm executes retransmission operation. So it is very important to decrease retransmission count. The BP period, called back off period, is equal to 80 bits or 0.32ms. The BP must align with the beginning of each super frame. All CSMA/CA operation must align also with BP period. Prior to transmission the node must ensure if there is sufficient time left in the CAP for the transmission and any consecutive acknowledgments packets, if positive feedback option is selected. As a result of the implementation, back off delay is decreased and potential packet collisions are reduced. The result indicates the improvement in network traffic performances and shows hardware performances. Four sensor nodes are taken for analysis of MAC protocols. A network coordinator is connected to other three nodes through wireless channels at 2.4GHz. Other sensor nodes collect information and send it to network coordinator.

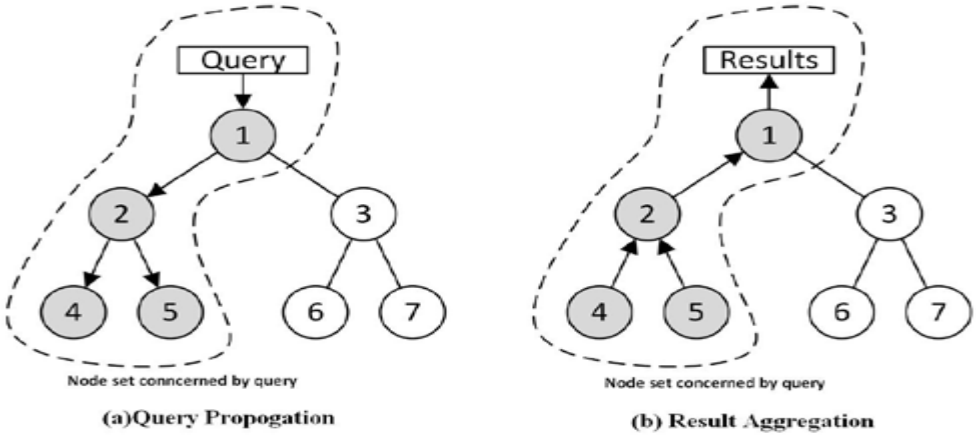

Fig .4 Query Propagation and Result Aggregation

4.2 VHDL Modeling of CSMA/CA Protocol

The VHDL (Very High Speed Hardware Description Language) is defined in IEEE as a tool of creation of electronics system because it supports the development, verification, synthesis and testing of hardware design, the communication of hardware design data, and the maintenance, modification, and procurement of hardware. It is a common language used for electronics design and development prototyping. After simulation analysis, hardware characteristics can be extracted from synthesis operation. To make things easier a modeling phase is first handled before developing algorithm describing

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

system functionalities. Modeling is an abstract representation of a physical reality. System models are created to be able to reason about certain properties of the system's behavior and to serve as a specification for the design process that will lead to a physical implementation of the system, which is compliant with the model.

CSMA/CA module needs a waiting state when executing the protocol. It uses counters in various times for channel access control. Particularly, a node has to wait for random back off duration before attempting carrier sensing and it has to wait for contention window slots for clear channel assessment (CCA). The functionality of counter is then implemented in this system. The waiting time is counted with respect to the clock speed used. CSMA/CA module interacts with a number of functional modules. Firstly, CSMA/CA module is a model of a protocol interactively related with the physical layer of the wireless devices. Thus, the CSMA/CA block is connected to an extra physique block in order to provide proper physique response related to the request. This extended module is useful to control the channel state. Secondly, the CSMA/CA system needs a Random Generator module.

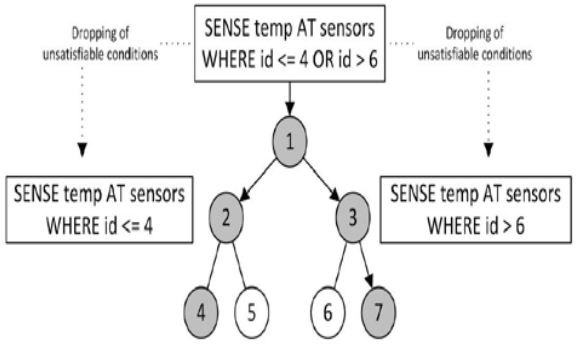

Fig.5 Dropping of unsatisfiable condition parts in-network

5.CONCLUSION

We introduced research circumstances of big data in the field of sensing. We first introduce different applications that deal with big sensing data and then summarize techniques used to solve the big sensing data problems. Finally, we propose some future research directions. A large number of platforms which have the capacity for sensing at the city level are still in the designing concept stage, but a lot of research methods have been proposed. Though most of them are based on existing data processing and management techniques, they are still very useful. Mobile sensing and Smartphone applications are still considered as the most popular topic. Researchers will dedicate themselves to Smartphone applications in the near future because it is the most mature large-scale sensor network so far. Our clustering method is based upon a modified Expectation Maximization technique.

REFERENCES

[1] X. Wu, X. Zhu, G. Wu, and W. Ding, "Data mining with big data,"Transactions on Knowledge and Data Engineering, vol. 99, June 2013.

[2] G. Jung, N. Gnanasambandam, and T. Mukherjee, "Synchronous parallel processing of big-data analytics services to optimize performance in federated clouds," IEEE Computer Society, 2012, pp. 811–818.

[3] Y. Demchenko, Z. Zhiming, P. Grosso, A. Wibisono, and C. de Laat, "Addressing big data challenges for scientific data infrastructure," IEEE Computer Society, 2012, pp. 614–617.

[4] M. Rezaei, M. Sarshar, and M. M. Sanaatiyan, "Toward next generation of driver assistance systems: A multimodal sensor-based platform," vol. 4, February 2010, pp. 62 – 67.

[5] S. Blokzyl, M. Vodel, and W. Hardt, "A hardware accelerated real-time image processing concept for high-resolution eo sensors," September 2012. [Online]. Available: http://d-nb.info/1028174624

[6] K. Fall, "A delay-tolerant network architecture for challenged internets,"(SIGCOMM). ACM, 2003, pp. 27–34.

[7] M. Vodel and W. Hardt, "Data aggregation in resource-limited wireless communication environments - differences between theory and praxis,"(ICCAIS2012). Ho Chi Minh City, Vietnam: IEEE Computer Society, November 2012, pp. 282–287.

[8] M. Vodel and W. Hardt, "Data aggregation and data fusion techniques in topologies - a critical discussion,"TENCON 2012. IEEE Computer Society, November 2012, pp. 1–6.

[9] D. Laney, The Importance of 'Big Data': A Definition. Gartner, 2012.

[10] S. Madden, R. Szewczyk, M. J. Franklin, and D. Culler, "Supporting Aggregate Queries Over Ad-Hoc Wireless Sensor Networks," in Proceedings Fourth IEEE Workshop on Mobile Computing Systems and Applications, 2002, pp. 49–58.