Challenges to using decentralized spatial algorithms in the field: The RISERnet geosensor network case study

Matt Duckham, Xu Zhong, Kevin Toohey
University of Melbourne, Australia

Abstract

Over recent years considerable research effort has focused on developing decentralized algorithms for highly distributed computing environments, such as wireless geosensor networks. There are several putative advantages of decentralization, including scalability, energy efficiency, and operational latency. However, decentralized algorithms today are primarily found in simulations and lab-based deployments, but rarely if ever in the field in true deployments. In this paper, we review the principles of and drivers behind decentralization. We then contrast these drivers with a recent field deployment of a large wireless geosensor network for monitoring environmental conditions relevant to wildfire hazard, called RISERnet. The comparison highlights the key areas of difference, where current technology and applications of wireless geosensor networks are not yet able to take advantage of decentralization.

1 Decentralized algorithms

A decentralized system is a special case of a distributed system where no single system component knows the entire system state [10]. Decentralized systems have a long history in computer science. Important fundamental advances have been made in decentralized algorithms, such as in efficient algorithms for *leader election* [12]. The design of decentralized *spatial* algorithms have a more recent history, but have grown into an active area of current research [11].

Decentralization is especially well-adapted to highly distributed systems, such as wireless sensor or geosensor networks. In a decentralized wireless geosensor network, computing happens in the network itself, with decentralized algorithms run in parallel on every node. There are four primary potential advantages of using decentralized algorithms, when compared with centralized alternatives [5]:

1. *Scalability*: As networks scale from tens to hundreds, thousands, or even millions of nodes, centralized architectures struggle to manage the increasing number and complexity of connections. Decentralized architectures are more scalable, allowing nodes to be easily added or removed, with each node executing the same procedures in parallel with its peers.

2. *Energy efficiency*: Decentralized algorithms enable in-network computing, where low-level data processing and filtering can occur in situ amongst groups of nearby node neighbors. In turn, this can lead to less information being communicated between nodes, improving the overall energy budget of the network.

3. *Operational latency*: Processing data in the network may avoid the need to communicate data from the network to a central sink for processing. For sensor-actuator networks, where data may be needed in the network to effect changes using actuators, reducing communication can in turn reduce the operational latency of the system.

4. *Managing information overload*: The volume of data generated even by relatively small wireless geosensor networks can be formidable, with each *individual* data item typically of almost no value. Enabling low-level collaborative processing and filtering of data in the network embeds intelligence in the network, reducing the potential for overwhelming applications and users with near-meaningless data.

In the domain of geographic information and wireless geosensor networks (GSNs), these advantages have fueled considerable research interest in the design and development of decentralized spatial algorithms. Examples include [8], who examined efficient spatial queries in the context of geosensor networks; and [13], who developed a decentralized sweep algorithm for GSNs, used as a building block for many higher-level algorithms. Many examples of other decentralized spatial algorithms, such as sweeps, can be found in [5].

Despite these advantages, it is fair to say that relatively few decentralized spatial algorithms have been tried and tested in the field. In the following section 2 we look at a specific example of a wireless geosensor network, called RISERnet. Despite the putative advantages of decentralization, discussed above, RISERnet currently does not make significant use of any decentralized algorithms. Hence, Section 3 examines more closely the reasons for this omission in the case of RISERnet, with lessons for GSN and decentralized spatial algorithms more generally.

2 RISERnet: A GSN in the field

The RISER project (resilient information systems for emergency response) is developing technologies capable of capturing, collating, and communicating timely and relevant information, even in the extreme and unexpected circumstances surrounding an emergency. As part of the RISER project, a GSN for monitoring environmental variables relevant to wildfires has been developed and deployed in the field. Wireless geosensor networks are recognized as an important tool for environmental monitoring for wildfire applications [1], with past work including both image-based monitoring of forests [6, 9] as well as fine-grained monitoring of environmental conditions relevant to wild fires [2].

The overall architecture of the RISER system is shown in Figure 1. This architecture includes three main components: the GSN itself (RISERnet); a stream-processing middle tier, including database archiving; and a real-time user interface.

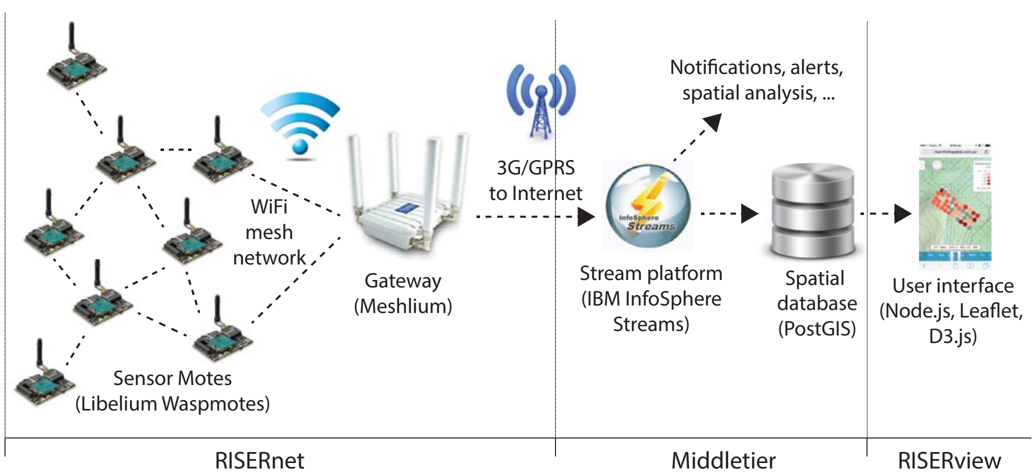


Figure 1: RISER system architecture, after [15].

2.1 RISERnet

RISERnet itself is a redeployable wireless sensor network for measuring environmental conditions pertinent to wildfire hazard and behavior. The network currently consists of 70 wireless sensor motes (Libelium Waspmotes). Each geosensor node in the RISERnet network captures data from its on-board sensors a regular interval (configurable, but by default every 60 minutes). Nodes communicate real-time sensed data to one of four special gateway nodes (called Meshliums) via an XBee DigiMesh (2.4 GHz) multihop network. Gateway nodes forward aggregated data via a 3G WAN connection to the middle-tier.

Mesh networking The multihop DigiMesh mesh network protocol eliminates the need for always-on coordinators. This reduces the frequency of battery changes in field. Furthermore, as the network does not rely on coordinators, the network is more resilient to single node and link failures. Although the line-in-sight communication range of the XBee module can be hundreds of meters, undergrowth and trees cause significant signal attenuation. In practice, the motes are on average about 80m from their nearest neighbor. Different network topologies have been adopted across the three different deployments where RISERnet has been used. The network has been deployed in a regular grid as well as an irregular deployment, to adapt to changes in vegetation density (motes were distributed more densely in the denser-vegetation, where signal attenuation is higher; in sparse-vegetation, the motes were positioned farther from each other).

Figure 2: RISERnet wireless sensor mote deployed in Olinda, Australia.

Time synchronization Power constraints exist in most wireless sensor networks deployments. Minimizing power consumption is critical for wireless sensor networks in the forest because a) higher signal attenuation leads to more frequent communication failures and retries; and b) solar panels are not usable in the forest. The most efficient solution for RISERnet was found to be duty cycling, where the network hibernates between sampling cycles. However, in order to establish multihop communication, all the motes must wake up simultaneously. Thus, time synchronization is essential in RISERnet. As the spatial extent of the network is relatively small (less than one square kilometer), RISERnet uses a simple time synchronization protocol. At the beginning of a sampling cycle, the gateway broadcasts a beacon which contains the current system time of the gateway and the next wake-up time of the network. The motes then synchronize their clocks with the gateway and set the wake-up time to that contained in the beacon message. The sampling period of the networks can then be changed simply by adjusting the setting in the gateway, which can be achieved remotely from the middle-tier.

Deployments The first RISERnet network was deployed in Olinda in Victoria, Australia (see Figure 2). Each mote was armed with a temperature and relative humidity sensor (SHT75), a soil moisture probe (Watermark), a weather station probe, and a solar radiation sensor (SQ-110). Across the area covered by the network, these sensors enable the monitoring of factors relevant to wildfire hazard, such as fuel moisture content (FMC), at fine spatial and temporal scales.

A significant feature of RISERnet is its extensibility: the network can easily be redeployed, and reconfigured with new sensors. For example, a new RISERnet network armed with two new types of sensors has recently been deployed in Powelltown, Victoria. The new sensors, originally designed to measure soil temperature (a PT1000 sensor) and soil moisture (a VH400 sensor), are used to monitor fuel conditions directly.

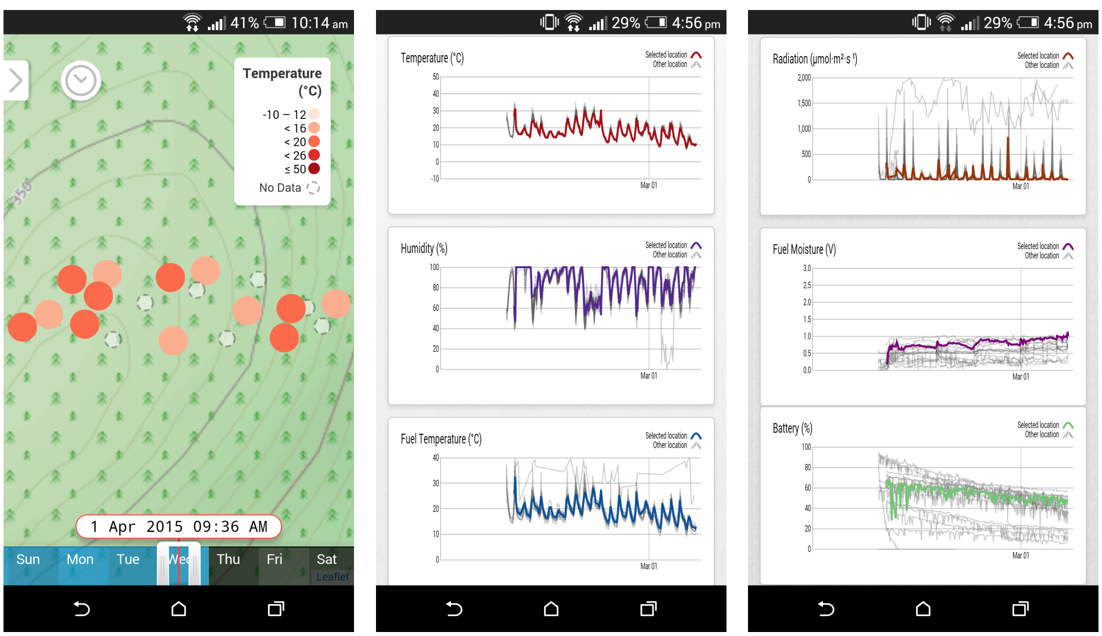
Indeed, redeployment is now simple enough for a third small RISERnet network to have been established in Anglesea, Victoria, by local primary school children, as part of a broader and longer-term fire education initiative involving the whole Anglesea community. With assistance from the RISER team, the deployment was planned and conducted by school children. The children also maintained the network, replacing batteries as necessary and fixing any problems with the sensors that occurred following deployment.

2.2 Middle-tier

The gateways of the RISERnet networks forward the data collected by the sensor motes to the middle-tier through a messaging protocol called MQTT [7] over a 3G mobile network. Such real-time data is more suited to online data management systems than offline systems. Stream processing systems are amongst the most familiar class of information systems that adopt an online approach to information processing. RISER uses IBM InfoSphere Streams, a commercial stream computing platform with a modular, component-based programming model. The RISERnet stream processing system forms a bridge between the sensor data streams emanating from the RISERnet network and all the uses of these streams. These uses include archiving in a traditional spatial database (MySQL); processing to generate interpolated, high-resolution maps of current conditions; and display of data in a real-time user interface.

2.3 Interface

The RISERview interface presents current and historical data from RISERnet. The RISERview interface shown in Figure 3 is built on top of a Node.js server using a Leaflet and D3.js based map interface. The interface includes a timeslider to allow access to both real-time and historical sensor data (Figure 3(a)) over the whole network, as well access to more detailed current and historical data from individual nodes (Figures 3(b) and (c)).



(a) Distribution of wireless sensor motes in Powelltown

(b) Plots of temperature, relative humidity, and fuel temperature

(c) Plots of solar radiation, fuel moisture, and battery level

Figure 3: RISERview mobile phone interface

3 Comparison

The RISERnet system described above currently operates without any use of sophisticated decentralized algorithms, despite the RISER team possessing significant expertise in their design and implementation. An obvious question is to ask why this is, and whether this omission arises from a failure in the sensor network technology, in the decentralized algorithm research, or some combination. This section looks briefly at each of the potential advantages of decentralization in the context of the RISERnet sensor network.

3.1 Scalability

Scalability is perhaps the most important potential advantage of decentralization. Even though the RISERnet network is a modest 70 nodes in size, the network is significantly larger than previous GSN deployments in forest environments (e.g., compare with [3, 4, 6, 9, 14], all less than 35 node networks) and scalability remains a challenge for RISERnet. Although the network is designed to be redeployable, for example, it does still require significant human labor to physically uninstall, move, and reinstall the network, with every manual operation repeated 70 times over, at significantly different locations, and in remote and harsh natural environments. Even after installation, the network is vulnerable to animals, such as small mammals. For example, a damaged weather station probe and a damaged sensor cord are shown in Figures 4(a) and (b), respectively. Cables that run along the ground are housed inside a garden hose and are protected at the ends with tape and rubberized fabric.

(a) Damaged weather station probe (b) Damaged sensor cord

Figure 4: Examples of a damaged anemometer (a) and a damaged sensor cord (b).

While decentralized algorithms might be able to offer only limited assistance with such physical scalability issues, a purely peer-to-peer approach does hold the potential to improve operating reliability and simplify the networking and (re)deployment process.

However, in practice such scalability advantages are difficult to realize. Most significantly, the reliability of the underlying mesh networking services of today's sensor network technology is not high enough to support reliable decentralized algorithms. The XBee DigiMesh multihop wireless communication protocol automatically finds a route (if there is one) to relay a message to the gateway, even when node or link failures happen. Hence, in most cases data from most nodes can be relayed to the gateway despite the relatively high frequency of node and link failures that occur even in relatively simple RISERnet network topologies and small network diameters (of at most ten nodes) used in our deployments. However, decentralized algorithms require not only reliable multihop communication between nodes and the gateway, but between all nodes in the network. Such reliable peer-to-peer networking is a prerequisite of decentralized computing, and is not well supported by current networking technologies.

Second, decentralized algorithms remain challenging to code and debug. One example of an early protocol failure in RISERnet occurred when a gateway fault resulted in repeated and frequent "handshake" messages from motes. This in turn led to rapid depletion of the mote batteries, leading to a loss of data and necessitating the replacement of all the batteries in the network several weeks earlier than originally scheduled. The high practical and labor costs of such errors tend to militate against using anything but the most basic in-network protocols.

Finally, although 70 nodes deployed in the field is a moderately large research network, RISERnet is not large enough to provide the levels of spatial detail required by many spatial algorithms. Hundreds or thousands of nodes in a single area may be needed to provide a fine-enough spatial granularity to discern complex spatial events, such as topological changes in regions. Today's wireless sensor network technology is not yet at the level where such large networks are practical or affordable, and so the need for decentralized spatial algorithms is lessened.

3.2 Energy efficiency

Developing an effective energy budget that enables the network to continue operating for extended periods of time remains one of the most challenging aspects of deploying any wireless geosensor network today. Extensive work on RISERnet has led to a network that can operate without battery replacement for up to six weeks. Achieving this longevity has required a range of technical innovations, including the modification of nodes to enable external rechargeable battery packs to be used (Figure 5). The gateways do not support automatic hibernation and wake-up, and so customized power controllers needed to be developed to turn on and off the gateway automatically, making sure to shut down the gateway's file system before powering it off.

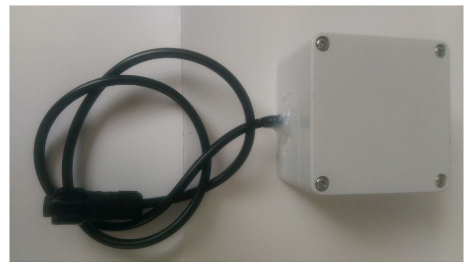


Figure 5: A customized external battery pack for a RISERnet node.

As discussed above, duty cycling is used to ensure the network can be powered off for any times when the network is not capturing data. Given our application of monitoring environmental parameters relevant to wildfires, a duty cycle frequency of between 15 minutes to an hour was adequate to capture any salient changes to the environment. In the context of such duty cycling frequencies, the costs of data communication are relatively small components of the energy budget, when compared to the costs of powering up nodes and sensors and the communication overheads required for time synchronization. In this context, decentralized spatial algorithms have the potential to offer only marginal benefits to the overall energy budget.

3.3 Operational latency

The operational latency of RISERnet (the time between data being captured and being made available to an end user) is low compared with the RISERnet duty cycle frequency. While our application needs can be satisfied by sensing data every 15–60 minutes, using our stream-based architecture in Figure 1 the sensed data is available online within a few minutes of being captured. Since these operational latencies are significantly shorter than the duty cycle periodicity, there is little to be gained from achieving the lower operational latencies that might result from decentralized algorithms. It is conceivable that other applications (such as monitoring of active firefronts) might indeed benefit from operational latencies lower than a few minutes. However, in the context of ongoing environmental monitoring, a centralized architecture is typically adequate.

3.4 Information overload

Information overload is an important issue, even for wireless geosensor networks of modest size and sensing frequencies such as RISERnet. The data generated by only 70 nodes, sampling a few times an hour for three months (the current length of time RISERnet has been continually operating) can still give rise to information overload issues. Note that it is not the *volume* of data that poses the problem; such data volumes and frequent updates are not especially challenging on their own. Rather it is the volume of data given the value of each individual data item. For example, knowing that node W08 measured a fuel temperature of 13.57°C at 13:41 on April 05 2015 is in isolation practically meaningless. Rather, it is only information about the upward trend of fuel temperature across a nearby region of nodes over a period of days, in combination with a coincident drop in fuel moisture, that together gives rise to meaningful knowledge (about potentially increased fire hazard). Thus, managing the potential for information overload concerns managing the volume of data in the context of the meaningfulness of individual data items.

In the case of RISERnet, decentralization certainly might assist with managing information overload, by identifying meaningful coordinated changes in the network. However, as decentralization is already of only marginal benefit (as a consequence of other scalability, energy, and latency considerations), the stream processing architecture used in RISERnet is also capable of managing the potential for information overload. Stream operators executed over the RISERnet data are able to identify salient and coordinated changes in sensed data. In short, decentralization is one mechanism for managing the potential for information overload, but not the only mechanism available to the GSN system architect.

4 Conclusions

This paper examined that gap that still exists between research into decentralized spatial algorithm design and practical wireless geosensor network deployments. With today's technology, practical considerations often negate the potential advantages of decentralized algorithms. Specifically,

- The low reliability of truly peer-to-peer mesh networking combined with the low levels of spatial detail afforded by networks comprising dozens rather than thousands of nodes leads to difficulties realizing the potential for increased scalability of decentralization.

- The impact of duty cycling upon the energy budget of the network may dwarf the costs of communicating data to a central gateway, particularly in cases where relatively low-frequency updates are required, such as ongoing environmental monitoring.

- In applications, such as monitoring ongoing environmental changes, where operational latency is lower than the required sensing frequency, decentralization has the potential to offer only marginal benefits in terms of operational latency.

- While the potential for information overload remains an issue even for networks of modest size, decentralization is not the only approach capable of helping to manage large volumes of volatile, low-reliability data. Hence, in networks where decentralization is not required for one or more of the reasons above, other techniques such as stream processing can substitute well in identifying meaningful patterns in dense data.

In summary, while clear reasons for the interest in decentralization exist, today's technology combined with some application requirements may not always require decentralization. However, as the technology improves (e.g., increased ease of embedded programming, lower hardware and deployment costs of sensor nodes, increased reliability of peer-to-peer mesh networking), the potential benefits of using decentralization may also increase. Further, with improvements in technology, it is to be expected that new applications with requirements for higher frequency monitoring will become more commonplace, also strengthening the case for decentralization. Thus, while decentralization is not necessarily the approach for now, it still seems plausible it may be a significant component of the approach of the future.

References

[1] A. Alkhatib. A review on forest fire detection techniques. *International Journal of Distributed Sensor Networks*, 2014, 2014.

[2] Y. E. Aslan, I. Korpeoglu, and Ö. Ulusoy. A framework for use of wireless sensor networks in forest fire detection and monitoring. *Computers, Environment and Urban Systems*, 36(6):614–625, 2012.

[3] K. Bouabdellah, H. Noureddine, and S. Larbi. Using wireless sensor networks for reliable forest fires detection. *Procedia Computer Science*, 19:794–801, 2013.

[4] D. M. Doolin and N. Sitar. Wireless sensors for wildfire monitoring. In *Smart Structures and Materials*, pages 477–484, 2005.

[5] M. Duckham. *Decentralized Spatial Computing: Foundations of Geosensor Networks*. Springer, Berlin, 2014.

[6] C. Hartung, R. Han, C. Seielstad, and S. Holbrook. FireWxNet: A multi-tiered portable wireless system for monitoring weather conditions in wildland fire environments. In *Proc. 4th International Conference on Mobile Systems, Applications, and Services*, pages 28–41. ACM, 2006.

[7] U. Hunkeler, H. L. Truong, and A. Stanford-Clark. Mqtt-s—a publish/subscribe protocol for wireless sensor networks. In *Communication systems software and middleware and workshops, 2008. comsware 2008. 3rd international conference on*, pages 791–798. IEEE, 2008.

[8] G. Jin and S. Nittel. Towards spatial window queries over continuous phenomena in sensor networks. *Parallel and Distributed Systems, IEEE Transactions on*, 19(4):559–571, 2008.

[9] J. Lloret, M. Garcia, D. Bri, and S. Sendra. A wireless sensor network deployment for rural and forest fire detection and verification. *Sensors*, 9(11):8722–8747, 2009.

[10] N. A. Lynch. *Distributed Algorithms*. Morgan Kaufmann, San Francisco, CA, 1996.

[11] S. Nittel. A survey of geosensor networks: Advances in dynamic environmental monitoring. *Sensors*, 9(7):5664–5678, 2009.

[12] N. Santoro. *Design and Analysis of Distributed Algorithms*. John Wiley & Sons, Inc., New Jersey, 2007.

[13] P. Skraba, Q. Fang, A. Nguyen, and L. Guibas. Sweeps over wireless sensor networks. In *Proc. 5th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 143–151, New York, NY, USA, 2006. ACM.

[14] G. Tolle, J. Polastre, R. Szewczyk, D. Culler, N. Turner, K. Tu, S. Burgess, T. Dawson, P. Buonadonna, D. Gay, et al. A macroscope in the redwoods. In *Proc. 3rd International Conference on Embedded Networked Sensor Systems*, pages 51–63. ACM, 2005.

[15] X. Zhong, A. Kealy, G. Sharon, and M. Duckham. Spatial interpolation of streaming geosensor network data in the riser system. In *Web and Wireless Geographical Information Systems 2015*, volume 9080, 2015.