# A Case Study to Monitor Cane Toads in Kakadu National Park

Saurabh Shukla, Nirupma Bulusu, Sanjay Jha
Networking Group
School of Computer Science and Engineering
The University of New South Wales, 2052 , Australia
E-mail: sshu495@cse.unsw.edu.au, nbulusu@cse.unsw.edu.au, sjha@cse.unsw.edu.au

UNSW-CSE-TR-0402
January 2004

THE UNIVERSITY OF
NEW SOUTH WALES

School of Computer Science and Engineering
The University of New South Wales
Sydney 2052 Australia

**Abstract**

Recent advances in wireless communication have promoted the large-scale research in development and deployment of sensor networks. Networked sensors that coordinate among themselves to perform large tasks are expected to revolutionize information gathering and processing in near future.

This thesis addresses the problem of large scale sensor deployment using the application of monitoring Cane toads in Kakadu National Park as a case study. Cane toads were introduced in Australia by mistake in 1935. Their uncanny ability to survive in diverse climates and lack of natural predators in the Australian ecosystem have promoted unhindered growth of Cane toads for the last 68 years This application is of tremendous importance to Australia because cane toads are endangering native species and the ecosystem.

A study of deployment requirements is important because it influences the network and system architecture; and consequently the design considerations for higher layer protocols and algorithms.

Previously proposed deployment work (especially in the sensor networks context) tries to achieve a single objective (e.g.., maximizing sensor coverage in a given area; or maintaining radio connectivity.) Deployment has not really been studied in terms of a higher-level application perspective when many objectives have to be satisfied simultaneously.

This work bridges that gap. Our thesis is that deployment is really a multi-variate problem and we provide a novel framework to studying deployment by integrating application, economic, and networking/technology objectives.

Specifically, the contributions of the thesis are:

- A framework in which the deployment problem can be reduced to:
  - Zone division: Division of deployment area into zones.
  - Zone classification: Classification of zones based on deployment priorities.
  - In-zone deployment: Strategies for deploying nodes within a zone to meet the bandwidth and coverage requirements.

- Observation that it is hard to get initial deployment right due to uncertainty. Bayesian framework is used for addressing uncertainty in domain knowledge and using it to drive adaptive learning algorithm
- Discussion of evaluation strategies

Working through a deployment strategy for cane toad monitoring reflects a hierarchical hybrid network of possibly many mutually disconnected clusters; which is counter-intuitive to the large-scale "flat" network models commonly assumed. Although our study is in the context of a single specific application, we hope the insights from our study will be useful to designers and researchers in the area of sensor networks.

The final aim is to assist the ecologist and biologist in their pursuit of limiting the growth of toads in the region. The goal is to develop a deployment strategy for sensor networks in Kakadu National Park. The sensor network thus designed will be used to monitor and track the presence of cane toads in the Kakadu National Park.

# 1 Introduction

## 1.1 Motivation

Wireless Sensor Networks (WSNs) are increasingly dominating the global research agenda. It is one of the 10 emerging areas of technology identified by MIT's Technology Review, which will soon have a profound impact on the economy and how we live and work. In the August issue of Business Week [8], Robert D Hof did a search for the next big thing. One of the next big things he came up with was the technology that lets cheap smart tiny devices create their own network, allowing them to send data on small range. Large-scale research in the field of wireless sensor networks has triggered the emergence of a wide range of applications where sensor networks will make a big impact.

Wireless Sensor Networks combine the challenges in distributed computing, network optimization and system energy optimization. Apart from looking into the use of sensor networks for a variety of applications, researchers are also working to devise routing protocols, distributed algorithms and techniques for building self-configuring and self-adaptable networks. Efficient power utilization figures is an important consideration in almost all the algorithms.

The application domains where these sensor networks could be deployed include industries such as transportation, health care, disaster recovery, military, security, habitat monitoring, and precision agriculture.

Sensor networks are made up of small devices called sensor nodes. The nodes combine a processor, memory, and sensors in a single chip to monitor parameters of interest, and have just enough radio power to transmit snippets of data to nearby devices to forward to other devices, thereby creating a network. Networking these small sensor nodes by radio links allows the sensor nodes to perform tasks that are hard to do using traditional sensors and traditional wired networks.

Mathematically, sensor networks can be looked up as an array or a set of highly distributed objects (sensor nodes). These objects coordinate amongst themselves to achieve a sensing task. The objects have the capability to sense and gather information on the parameter of interest within the area of investigation, perform computational tasks, receive and transmit information over wireless channels.

The sensor networks function unattended and therefore they need to be self-configuring and self-organizing. The nodes coordinate amongst themselves to attain the following two main objectives.

1. Sensing: Each sensor node acts like a source (of information) by collecting data through sensing and then extracting meaningful information out of it. The nodes perform one or many sensing tasks, which may include acoustic, temperature, humidity, pressure, movement, light etc. In the case where moving targets are involved, the nodes coordinate (using triangulation or similar techniques) to pin point the position of the moving target. The data collected through sensing may be processed locally at the sensor nodes to derive meaningful information or to compress it to reduce traffic over the network.

2. Networking: The collected data is routed via other sensor nodes to an external base station. A base station can be thought of as a gateway between the sensor network and the external world. Base stations are much bigger and powerful than a sensor node. The base station can be a static device, which receives data over wireless channels or a mobile robot that moves around in the area under investigation and collects information from the nodes. The nodes disseminate the collected data by passing it to a nearby node iteratively until the data reaches the base station thus creating a multi-hop wireless network. Nodes act independently to route data without needing a central authority.

## 1.2 Growth of Sensor Networks

Research in sensors networks is motivated by both technology push and application pull. The push is provided by the rapid progress in computation and communication technology as well as the emerging field of miniaturized low cost, MEMS based sensors. The pull is provided by the application where the physical world can be remotely observed and influenced using the wireless sensor network infrastructure.

### 1.2.1 The Technology Push

The push towards Sensor network has been mainly due the developments in MEMS and Embdedded Systems.

Micro electromechanical systems (MEMS) are made up of electro mechanical components using the techniques used in the semi conductor industries [47]. MEMS have earlier been used for safety devices in automobiles and ink jet printer nozzles. MEMS Technology has advanced quite a bit recently and electromechanical solutions are constantly migrating to the MEMS field in many areas., especially the ones related with sensor technologies.

Embedded systems consist of a computer hardware RAM, ROM, EPROM, Flash, NVRAM , microprocessor, IO interfaces, Interrupt controllers, and software designed to perform a dedicated function. Processor, software and required custom hardware on a single wafer are increasingly becoming common. This means a shift from traditional large boxes with separate processors, RAM, ROM and peripherals to a small device with all required hardware on a small circuit board. The falling cost of processors and related hardware has fueled growth in embedded systems. Embedded systems are characterized by:

- Dedicated controlling of a specific real time devise or a function
- Self Starting (human intervention is not required)
- Hardwired and embedded systems versions are functionally identical
- Self contained (contains non volatile operating systems)

Traditionally embedded systems have used in washing machines, cars, microwave ovens, audio and video equipment.

MEMS, Embedded system along with transducers would see all of the components needed for a sensor to function built right onto one highly integrated chip, thus drastically reducing size and cost.

Transducers are converters that can measure elementary physical quantities such as light, pressure, temperature, humidity, motion and convert them into electrical signals. All sensors are devices based on transducers. Additional functionality is required to provide compatibility of transducers with the microprocessor's interface.

It is envisioned that the process of miniaturization governed by growth in MEMS and Embedded Systems would see a one chip intelligent sensor device with transducers, processor, IO interfaces, RAM, ROM all embedded into one chip that can measure certain parameters in various environments. The signals generated by the transducers will be processed by one board microprocessor.

### 1.2.2 The Application Pull

Sensor networks have enormous application potential. Few of the applications include:

- Military and Defense Systems

  Wireless Sensing Networks can replace big and high cost traditional equipment with large arrays of distributed sensors for both security and surveillance applications.

  Wireless sensors packed with low powered video surveillance cameras and transceivers will be capable of gathering, processing, receiving and sending streaming video data and therefore can be used for large-scale surveillance of an entire region.

- Environment Monitoring

  Sensors could be used in deserts, plains, ocean beds, mountains, ecological regions to detect and monitor environmental changes.

- Habitat Monitoring

  Miniature sensor nodes can be deployed in dense clusters in an ecological area of interest. They can be used to track and monitor birds, insects, and small animals which can be identified using acoustic recognition techniques. They will provide us with information which is hard to obtain using traditional instruments which are big and it is not feasible to spread them all around in a region.

- Traffic

  Wireless sensors can be loaded with motion sensors to monitor vehicle traffic on a highway or in a congested part of a city and be used to provide real time traffic congestion to other vehicles. Alternatively they could be used in the parking lots to determine which spots are occupied and which are free.

- Disaster Recovery

  The sensors networks can assist in areas of possible natural disasters such as forest fires. They can raise alerts as soon as fire or the possibility of fire is detected. In all fires, early warnings are critical in preventing small harmless brush fires from growing beyond control. They will also be capable of providing fire fighters with information such is direction of spread, thermal gradient, and wind conditions which would help fire fighters respond quickly and efficiently to forest fires. In all fires, early warnings are critical in preventing small harmless brush fires from becoming monstrous. By deploying specialized wireless sensor nodes in strategically selected high-risk areas, the detection time for such disasters can be drastically reduced, increasing the likelihood of success in early extinguishing efforts. Also, the nodes are self-configuring and do not need constant monitoring.

Other applications where sensor networks can be used are health monitoring, process control and automation, inventory management, precision agriculture.

## 1.3 The Problem Description

It is hard to generalize application objectives over completely disparate application domains. We pick a particular application "*cane toad monitoring*".This is an application of tremendous importance to Australia because cane toads are endangering native species and has attracted widespread national attention. They are about to colonize The Kakadu National Park, which is one of world's heritage sites. The Kakadu National Park is rich in habitat and more than ninety percent of the park is habitable by the cane toads, which makes it natural to monitor them in this region. The


Figure 1: Current Distribution of Cane Toads

dark region in the Figure 1 shows the current distribution of cane toads in Australian Northern Territory. Figure 2 shows the regions close to Kakadu where the cane toads have been sighted (source: Dam et. al. [16]).

The objective of this Habitat Monitoring case study is to conceptualize the deployment of an array of inexpensive, lightweight sensors that are capable of acoustical observations to monitor, track and measure the impact of Cane Toads in Kakadu National Park. We try to answers questions such as what kind of sensor devices to use, how many sensors to deploy, the density of sensors with in a region, the geometrical orientation and organization of sensors. We try to meet the application objectives while minimizing the dollar costs of the deployment. The sensors will sense the presence of cane toads, interpret information, log data and later transmit core data to a central facility. The sensor nodes receive analog data in form of toad vocalizations, digitalize it, process the information, and report the results. The central facility receives this report that contains several environmental parameters across a range of temporal ecological gradients.

## 1.4 Significance

Researchers have proposed different network architectures for various applications.

The architectures have been proposed with potential applications in mind but they lack insight into environmental dynamics that come into play when a real specific application is considered. All of the approaches make assumptions that sensors have been deployed uniformly at random throughout the terrain of interest and that the network consists of a singular kind of device.

This work aims to bridge the gap between use of sensors in potential application vis-a-vis real existing practical application deployment strategy.

## 1.5 Approach and Contributions

Previously proposed deployment strategies especially in the sensor network context., try to address a single dimension such as delay-tolerance, maximizing the area of coverage in the network [41], maintain radio connectivity for data transfer, target location detection [42] and self-deployment strategy [43].

Deployment has not really been studied in terms of a higher-level application perspective when many objectives have to be satisfied simultaneously. When these objectives are studied independently; the solutions would be radically different.

More over the researchers need to consider an application for deployment. Few inherent application characteristics bring in additional assumptions into the sensor deployment strategy which makes it unsuitable for other applications. Our work bridges this gap. We provide a novel framework for studying deployment by integrating application, economic, and networking/technology objectives.

Specifically, the contributions of the thesis are:

1. A framework in which the deployment problem can be reduced to:

   (a) Zone Classification: Division of deployment area into zones and classification of the zones based on deployment priorities.

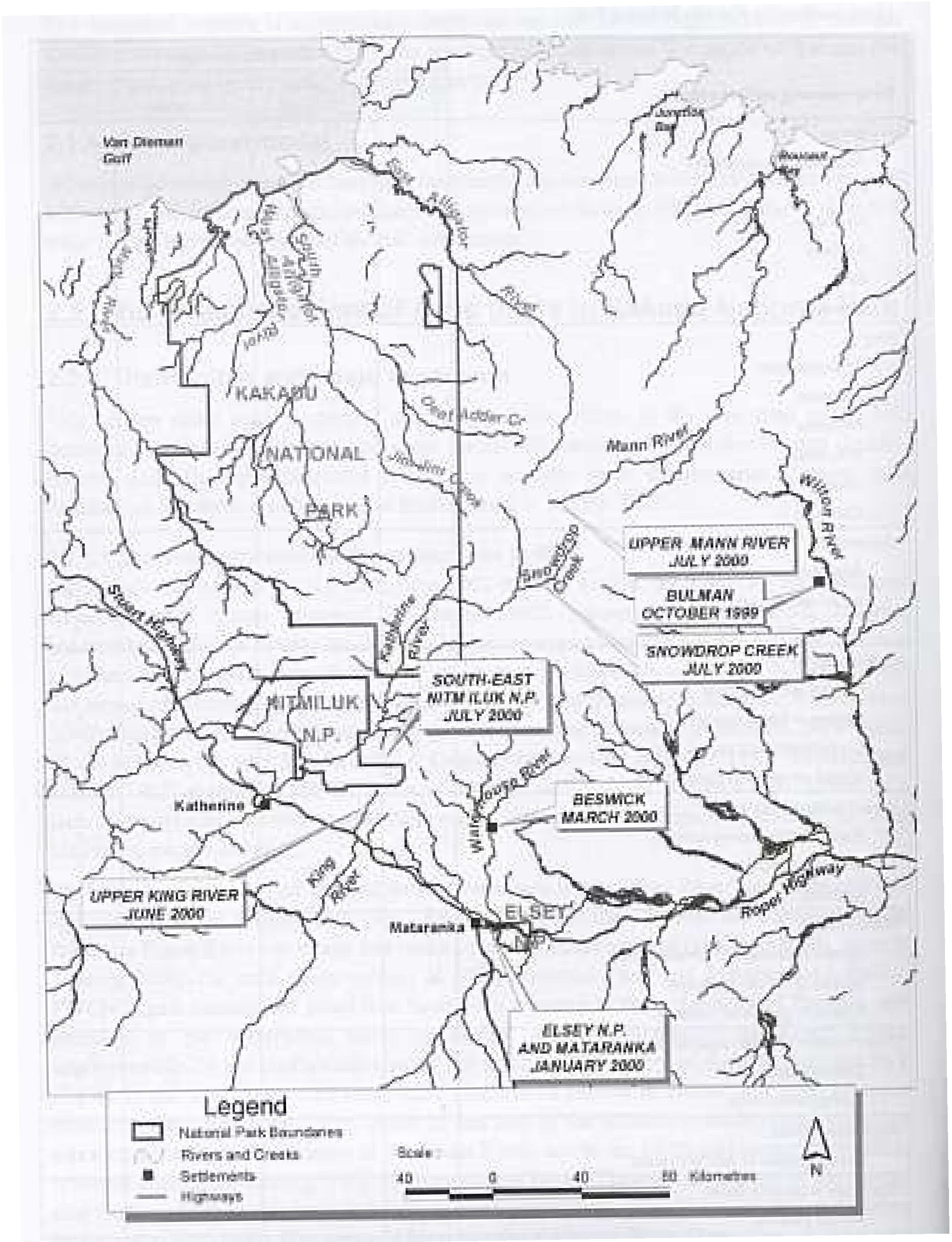


Figure 2: Current Status of Cane Toad

   (b) In-Zone Deployment: Strategies for deploying nodes within a zone to meet the bandwidth and coverage requirements.

2. Observation that it is hard to get initial deployment right due to uncertainty. Bayesian framework for categorizing uncertainty in domain knowledge and using it to drive adaptive learning algorithm

3. Discussion of evaluation strategies

Although our study is in the context of a single specific application, we hope the insights from our study will be useful to designers and researchers in the area sensor networks.

## 1.6 Thesis Organization

The rest of this thesis is organized as follows. Chapter 2 gives an overview of Wireless Sensor Networks and discusses background and related work. Chapter 3 describes at length the characteristics and behavior of the cane toads and the ecology of the Kakadu National Park. It also discusses state of art research work being done to halt the growth of cane toads and long-term research goals for this problem. Chapter 4 discuses the problem from the sensor deployment perspective, the design goals, the proposed solution, deployment strategy and the adaptive learning algorithm based on Bayesian Networks. Chapter 5 discusses the potential evaluation strategies and metrics to measure how successful sensor deployment has been. Chapter 6 concludes this work and discusses the scope of future work.

# 2 Background and Related Work

The previous chapter mentioned the use of sensor networks to monitor the growth of cane toads in the Kakadu National Park. In this chapter, we discuss work most closely related to ours in the area of sensor networks, hotspot and which we try to build upon deployment and machine learning.

## 2.1 Wireless Sensor Networks

### 2.1.1 An Overview

In recent times distributed ways of communicating, processing, sensing and computing are replacing more traditional centralized architectures. The projected trend is that the architectures will move away from single centralized highly reliable platform and move towards densely distributed cheap, lightweight and possibly unreliable individual components that are more capable of complex tasks as a group than one powerful computer.

Unlike traditional networks, the sensor networks consist of a large number of tiny autonomous sensor nodes. Every sensor node is mounted with sensors, data processing capabilities and wireless communications in the form of on board transducers, RAM, ROM, CPU, receiver, transmitter and a battery. Where deployed these sensors will be eyes, ears and noses of the system.

The sensor network's vision is to have thousands or millions of nodes networked together sensing their surroundings and coordinating amongst each other autonomously. Each sensor node acts as an information source, sensing and collecting data samples from its environment. Sensor nodes later disseminate the collected data, creating a multi-hop wireless network that routes data to other sensor nodes and to external destinations.

The following sections discuss the sensor network platforms, architectures, deployment characteristics of sensor networks, and the difference between traditional networks and sensor networks.

Figure 3: Intel Sensor Node

### 2.1.2 Sensor Platforms

The overall setup and configuration of the developed sensor node prototypes that have been developed and are currently used for experimental research are very similar.

Examples for such nodes include the Mica Motes developed by "University of California, Berkeley" and commercially manufactured by crossbow (Figure 6), the nodes (Figure 4) developed by the EYES project in European Union [48] involves "Technische University, Karlsruhe", "ETH, Zurich", "Freie University, Berlin", Intel [18] (Figure 3), MIT $\mu$Amps - Adaptive Multi Domain Power aware Sensors project [19] (Figure 5), i-Bean developed by 2 alumnus from MIT [20]. Millennial claims to have made the world's smallest existing ensor node.

A typical sensor node contains analog and digital interfaces, a radio transceiver, and a microcontroller for sensor signal processing, control and networking.

The microcontrollers used are: Atmel(Atmega 103, Atmega 128). Atmega 103 and Atmega 128 microcontrollers almost have identical architectures. The main difference is that Atmega 128 has a native two cycles multiply instruction which offers a big performance advantage.

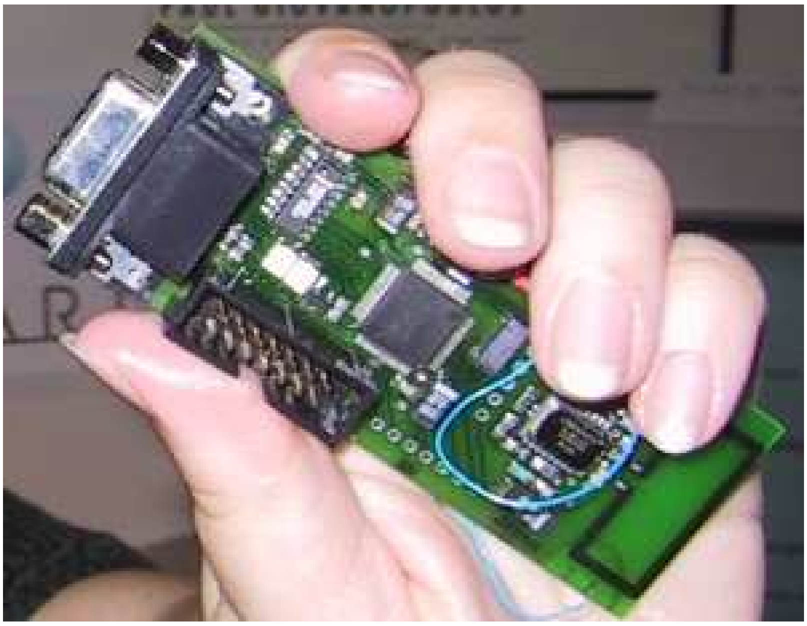

Figure 4: Eyes Sensor Node

Figure 5: $\mu$Amps

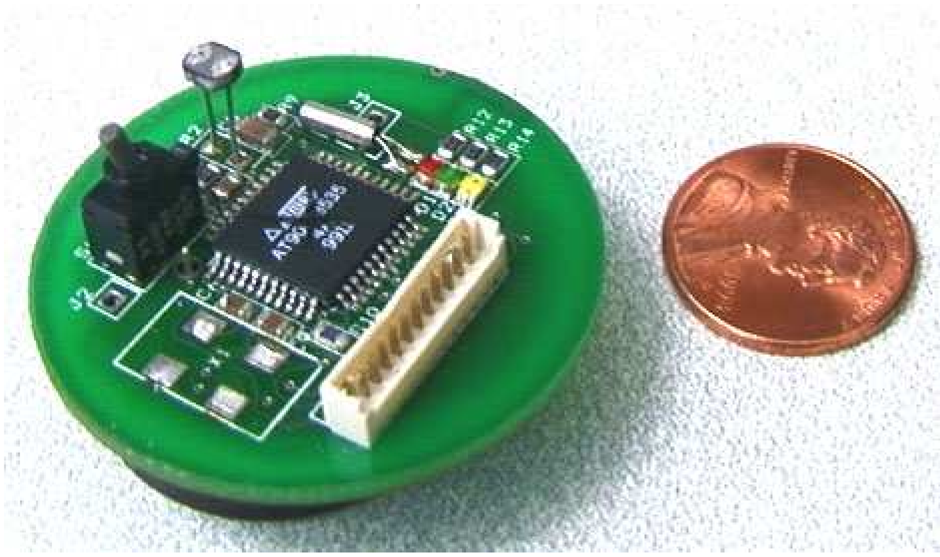

Figure 6: Berkeley Sensor Node

Motorola microcontrollers: The microcontrollers are 16MHz from the 68000 series. They also come along with acceleration and pressure sensor integrated circuits (ICs) . This will enable sensor networks with the convenience and reduced installation expense. This is achieved through their compatibility with the ZigBee protocol.

Texas Instruments(TI) MSP 430. Texas Instruments has created the MSP430 which enables system designers to simultaneously interface to analog signals, sensors and digital components. Texas Instruments has introduced four new additions to its line of MSP430 16-bit RISC microcontrollers (MCUs) which can provide 50 times more processing performance while maintaining low power consumption. The flash MCU feature improves the signal processing power. It comes with an integrated 10-bit analog-to-digital converter and a small form factor optimizing new devices for embedded applications. [21]

The radio modems used include those by RFM, Chipcon, Maxim, Nordic, RFMD, Xemics, Infineon [23].

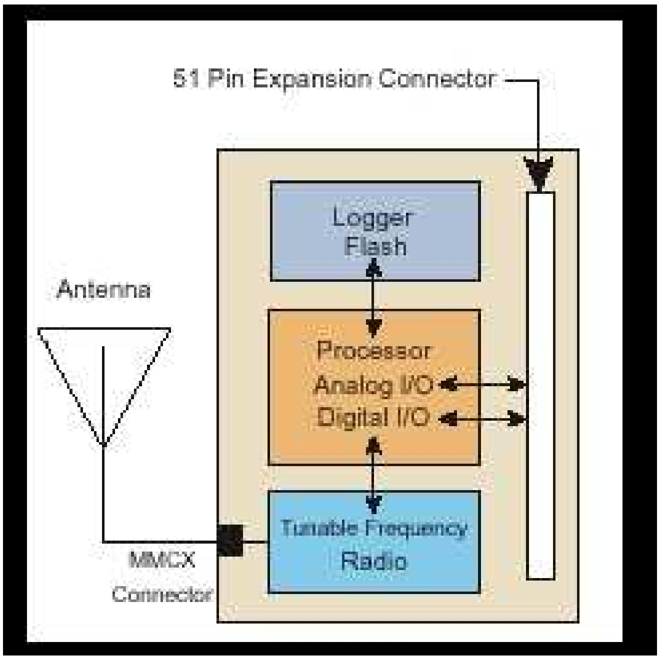


Figure 7: Block Diagram of Sensor Node

A typical sensor node architecture uses a DC booster to provide stable voltage from degrading alkaline batteries. Under zero load, a booster may draw between 200 and 300 $\mu$A, depending on the battery voltage. [22]. Figure 7 shows the block diagram of the mica sensor node.

When deployed commercially, the sensor devices will tend to be application specific, rather than general purpose, and carry only the available hardware support actually needed for the application.

### 2.1.3 Network architectures and models

Sensor networks are essentially means to collect data pertaining to a phenomenon and relay information about the phenomenon to the neighbors iteratively until the information reaches a base station.

The infrastructure consists of a set of sensors nodes and base station, and additional hardware, which may be required by the application. The sensor nodes perform computing, networking and application specific sensing. A base station is a powerful machine and acts as a link between the sensor network and the external world. The infrastructure is characterized by sensing range and accuracy, area under investigation, density of sensor nodes, battery life, network longevity, transmission range, processing capabilities, memory size and node mobility.

The data sensed requires pre-processing before meaningful information can be extracted out of it. The data processing can be centralized or distributed.

A centralized approach sends data from each sensor to a processing center where data processing and dissemination are carried out. This approach is good when the number of sensor nodes is relatively small and the data to be transferred is also less. However, when the amount of sensor nodes goes to hundreds or thousands, the centralized approach will not scale.

A better alternative to this problem is the decentralized or distributed approach. Several efficient decentralized sensor network architectures have been proposed that use some kind of tiered architectures. The proposed approaches try to design a network structure such that it facilitates decentralized data dissemination.

Tiered Architectures. The sensor nodes are becoming lightweight, smaller, cheaper and more powerful. How ever these advances bring in a trade off and constraints to sensor nodes in the from of processing capacity and memory available. A effective design on sensor networks will have to be role based. A large number of small sensor nodes will be utilized for low level tasks like sensing and will be supported by a lesser number of relatively powerful nodes having more memory, processing capabilities and power availability to perform complex higher level tasks.

Conventional data transfer approaches like flooding waste communication and energy resources by sending redundant data throughout the network. Protocols need to be adaptive and aware of the available resources. A number of routing protocols like Directed Diffusion [25] and SPIN [28] have been proposed to address such issues [7].

- Directed Diffusion

  Developed at UCLA is built for information retrieval and data dissemination. It is task specific and provides good support for application where data is triggered by observation of a phenomenon. Directed Diffusion uses an Interest-Data handshake to disseminate data from source to sink.

- SPIN (Sensor Protocols for Information via Negotiation)

  Developed at MIT, SPIN has several similarities with directed diffusion but unlike directed diffusion the data transmission is triggered by the source and not the sink. It names its data using meta-data descriptors, similar to the interest attributes used in directed diffusion. The Meta data is mapped in a one to one relationship with the raw data. Like the interest in directed diffusion the format of meta-data is application specific.

**Comparison and Discussion**

SPIN and Directed Diffusion are both application specific and subject based protocols optimized for disseminating application specific data in a sensor network and try to achieve energy-efficient communication.

Both are data centric, which means all communications are based on the named data. In directed diffusion the data is named or identified by the attributes and in SPIN the data is named by Meta data. There is no network layer node identification used for routing. The application layer drives message exchange. The use of data naming allows negotiation between nodes concerning what data to forward to eliminate redundancy.

Both protocols are designed to be a reliable communication medium. The SPIN protocol is developed to overcome the classic flooding by negotiating between the nodes before the actual data is transmitted. Thus SPIN is sender initiated data communication protocol. In the case of directed diffusion, the sink and the source diffuse interests onto the network and if the interest matches, the source sends the data towards the sink. Therefore, in the directed diffusion protocol, the data communication is sink initiated.

Both protocols are energy conscious. In SPIN, the nodes poll the available resources and SPIN nodes can make decisions adaptive to the resource available. Similarly in directed diffusion, the Geographic and Energy Aware Routing (GEAR) helps in propagating interest messages to a specified geographical location defined by latitude and longitude, localizing the flooding in the network and helping to be more application specific.

In essence, both the protocols are flooding-based but vary from the classic flooding due to the implementation of different mechanisms to control the unwanted flooding in the network.

### 2.1.4 Communication Protocols

Communication protocols are important as they are the main consumer of power in a sensor. The lesser the communication overhead in the system, the longer is the network life. Hence, algorithms are needed for reducing the amount of communication in the network. The performance of the network protocol will be highly influenced by the network configuration dynamics, as well as by the specific data transfer model employed.

Data transfer can be categorized into network configuration and application data transfer. Network configuration refers to data transfer taken to establish and maintain the network. Application data transfer refers to transfer of raw sensed data or information inferred from sensed data.

### 2.1.5 Network Configuration

Network configuration protocols is responsible for tasks like establishment of network, creating and configuring paths to the base station, communication between sensor nodes and communication between base station and sensor nodes related to setting up of the network. It also includes tasks such as creating triggers to remotely wake up a sleeping sensor node upon observation of a phenomenon, recreating the path when a link is broken and recovering from unexpected system failures .

The routing approach can use one or more of the following approaches in tandem.

1. **Flooding or Broadcasting**: This is the simplest of cases. Each sensor nodes broadcast their information to their neighbors, which in turn broadcast the information to their neighbors until the information reaches the base station. This approach has a very high overhead and data redundancy, which is not highly desirable for a lot of applications. Data aggregation algorithms are being worked upon by researchers, which will reduce the overhead of broadcast [29]. One positive aspect of broadcasting is that flooding is not affected by the changes in topology of the network.

2. **Unicast:** The sensor nodes can send the data to a cluster head using one-to-one unicast which in turn sends it to the base station. The sensors can either communicate to the base station directly in one hop or using a multi hop via other sensor nodes.

3. **Multicast:** In a multicast approach, sensors form directed groups and use multicast to communicate among group members. The base station could communicate with any member of the group to obtain the desired data.

### 2.1.6 Application Data Transfer

The data gathered by observing a phenomenon by the sensor nodes is transferred along the paths established in network configuration.

Data transfer can be:

**Continuous**: The sensors continuously sense data and send it back to the base station. For example, simple temperature sensors in a critical environment such as a chemical plant need to sense the temperature at small intervals of time and send the data back to the base station.

**Triggered**: The data transfer can be triggered by observation of an event or phenomenon. In the case of bird monitoring a sensor node might send information as soon as the target bird is located. If the base station requires some information it will initiate a query and the sensors send back the gathered data. In this case the base station triggers the data transfer.

**Hybrid**: Data transfer can also be a hybrid of the triggered and continuous. Sensors send data after some unit intervals of time and can also send data open observation of a phenomenon.

As the cost of sensors decreases, it will increasingly affordable to deploy a large number of sensors in the field to the tune of hundreds of thousands. The sheer number of sensor node an application may require causes a scalability problem.

### 2.1.7 Characteristics of Sensor Networks

A few distinguishing characteristics of the sensor networks are:

- Sensor nodes are small light weight components with limited energy resources, limited computational capacity and memory.
- The sensor nodes and the over all network is highly prone to failure thus the topology is dynamic and changes very frequently in sensor networks.
- The sensor nodes have to use broadcast to establish paths.
- Paths cannot be pre-configured.
- The sensor networks are self-organizing and self re-configurable.
- The sensor nodes may not have a unique identification identifier like the IP address in the traditional networks.

- Sensor nodes are inexpensive and are deployed densely in large numbers.
- Concurrency intensive operation: Data will be simultaneously captured, manipulated, and transmitted over the network. Robust operations:
- The sensor nodes will be numerous, and will function unattended and the application will be functional a large percentage of time.

### 2.1.8 Difference from Traditional Networks

Compared to classic networks, sensor networks present a few differences. The overall usage scenario and the implications that this brings to the traffic and the interaction with the users is quite different from traditional networks.

The most significant difference between traditional distributed platforms and wireless sensor networks is that in wireless sensor networks there are strict limitations on energy, computation, storage, and bandwidth resources. For example, the wireless sensor node manufactured by crossbow (Figure )has a 4MHz, 8-bit, Atmel Atmega 128, CPU, 512 KB of Flash memory, 4Kb EEPROM, a 10 Kbps RFM radio, and has to operate on two AA batteries.

Unlike traditional wired network nodes, sensors nodes are lightweight, cheap, low-energy, and able to process and communicate among themselves.

Another major difference is the type of traffic that is produced, and the requirements about its delivery. Quality of Service (QoS) cannot be defined in classic terms such as network throughput or packet delay, QoS in sensor network is determined by how well the sensors satisfy the application requirements. The sensor networks are fragile and therefore reliability factors are more important then packet delay.

Traditional networks have devoted routers. In sensor networks nodes also act as a router, forwarding data packets for other nodes.

In traditional networks, users are connected to a computer terminal (or group of terminals) and require a service from another computer. The net-

Figure 8: Mica Sensor

work acts as a medium bringing the two parties together. Sensor networks on the other hand are independent entities and more distributed systems. As mentioned earlier, the nodes communicate amongst themselves to gather data and later disseminate the gathered data. They provide the end user with information that is constantly changing as the phenomenon under observation is changing.

## 2.2 Hotspot Deployment

Hotspots occur when there is a lot of shared access to one destination, or a combination of source-destination pairs that result in the same collision within the network (the latter is uncommon). Hotspots are areas of high traffic intensity in the network.

In chapter 4 we discuss a novel method of identifying hotspots where sensor networks will be deployed to monitor the presence of cane toads in Kakadu National Park.

Although we are not aware of any previous work that considers the specific problem described here, this work is influenced by the following mentioned research in deployment.

### 2.2.1 Incremental Sensor Deployment

Howard et. al.[24] proposed incremental deployment algorithm for mobile sensor networks. The algorithm deploys nodes one-at-a-time into an unknown environment, with each node making use of information gathered by previously deployed nodes to determine its target location.

The authors assume that the sensor nodes are mobile and need to be location aware. "Incremental sensor deployment" describes an incremental and greedy deployment algorithm for sensor networks. The algorithm deploys nodes one at a time into an unknown environment, with each node making use of information gathered by previously deployed nodes to determine its target location. The algorithm is designed to maximize network coverage whilst simultaneously ensuring that nodes retain line-of-sight with one another (this latter constraint arises from the need to localize the nodes. The nodes can localize themselves in a completely unknown environment by using other nodes as landmarks).

The series of simulation experiments performed by the authors serve to characterize the performance of the incremental deployment algorithm. Nodes are deployed one at a time, with each node making use of data gathered from previously deployed nodes to determine its optimal deployment location. The algorithm is greedy in the sense that it attempts to determine, for each node, the location that will produce the maximum increase in the network coverage area.

The algorithm iterates through four phases:

- Initialization

  The nodes can be in only one of the three stages: waiting, active or deployed. A waiting node is waiting to be deployed, a deployed node is already deployed and an active node is in a transition from 'waiting' to 'deployed'. Initially all the nodes are set to the waiting stage. Only one of the nodes act as an initiator node and is in the stage 'deployed'. This node is the starting point and initiates for the deployment process.

- Selection

  The selection phase attempts to maximize the coverage metric while simultaneously satisfying the visibility constraint. i.e. nodes are in the line of sight with one another. Since there is no way of determining the maximum coverage a priori and knowledge about the environment is incomplete, the algorithm relies on sensed data from deployed nodes using heuristics to guide the selection process.

  As a first step, sensor data from the deployed nodes are combined to form a map divided into cells. Each cell is assigned one of the three states free, occupied or unknown. A free cell contains no obstacles; occupied cell contains more than one obstacles, and unknown if the cell has no information about obstacles.

  Bayesian approach is used to determine the probability that each cell is occupied, and then this probability is used to determine the state of each cell. The inherent nature of Bayesian approach provides rules to change the existing assumptions or probabilities in light of new evidence. This ensures that the visibility constraint is satisfied. Next, the sensor nodes are placed at a free cell in the map. Flood-Fill algorithm is used to determine whether such a cell is reachable starting from the location of each deployed node in turn one by one.

  This defines a reachability space, the selection algorithm tries to deploy the node to a position such that they have a higher potential to cover the greatest area of the unknown space.

- Assignment

  The assignment phase assigns the newly selected cell to the first waiting node and the node's state is changed from waiting to active in the execution stage. If an obstruction is encountered the obstructed node is swapped with the node obstructing it.

- Execution

  Active nodes are deployed sequentially to their locations. The state of each node is changed from active to deployed.

The algorithm iterates through the selection, assignment and execution phases, terminating only when all nodes have been deployed.

### 2.2.2 Modeling Mobility for Wireless Local Area Networks

Moving between different locations, network providers, switching between Ethernet, wired LAN to wireless LAN is a hassle for many users of mobile devices. Earlier technologies did not provide for uninterrupted connectivity as a user switched location and changed network provider. For example, a user unplugged the computer from the wired local area network at work and switched to a wireless local area network.

A public hotspot or an IP Zone based on wireless local area network is a specific geographic location within which a user can wirelessly connect to the network services. Hotspots are often located in heavily populated places such as airports, train stations, libraries, marinas, conventions centers and hotels. The hotspot gives users access to the network of multiple operators through a single billing relationship at a location.

To do a performance evaluation of a protocol for an ad hoc network, the protocol should be tested under realistic conditions including, but not limited to, a sensible transmission range, limited buffer space for the storage of messages, representative data traffic models, and realistic movements of the mobile users (i.e. a mobility model).

Several mobility models have been proposed [1]:

1. Random Walk Mobility Model (including its many derivatives): A simple mobility model based on random directions and speeds.

2. Random Waypoint Mobility Model: A model that includes pause times between changes in destination and speed.

3. Random Direction Mobility Model: A model that forces Mobile Nodes to travel to the edge of the simulation area before changing direction and speed.

4. Gauss-Markov Mobility Model: A model that uses one tuning parameter to vary the degree of randomness in the mobility pattern.

5. City Section Mobility Model: A simulation area that represents streets within a city.

All of the above mentioned models are able to model mobility using randomization or probabilistic estimation but suffer from the lack of realistic mobility and workload models. In this paper, Matthias et. al. [30] have suggested proposed an analytical model, which is a hybrid of an empirical mobility model and a synthetic traffic model.

The authors assume a three-tier architecture. The mobile nodes or terminals are associated to wireless base stations, radio access point (RAP) represents the last hop of provision network. The function of the first tier can be described as radio access. The second tier is built by radio access routers (RAR), RAS and RAP have point to point connection via up-links. One more edge gateway routers (EGW) form the third tier of the architecture.

For Mobile communication the models can be broadly categorized into two:

a) **Microscopic**: Models describe the mobility behavior of the individuals and are often based on analytical description of user movements.

b) **Macroscopic**: Models describe the aggregated effects of mobility obtained using statistical data collection. For example, the transportation infrastructure is modeled and trips are planned from aggregated individual's activity patterns. The behavior of individuals and geometrical distribution of households is derived from census data.

This work deals with macroscopic model. Travel modeling approach and Census data are the building blocks of this model. The basic elements of travel modeling approach are trip and zone. Trip is defined as movement of a user from source to destination. Trip is used to model user behavior. The zone defines an area having socioeconomic characteristics with an attraction level. The authors co-relate these two to predict the network infrastructure.

1. **Trip Generation**: The trip generation estimates the total number of trips that depart and arrive in a specific zone. Trips are classified by purpose; such is work, shopping, school, recreation. Trips help in estimating user densities for a given zone.

2. **Location classification**: The classification of location reflects their different attraction levels over time. A reasonable granularity can be reached by assigning a base attraction level to a zone. Zone can be classified into residential, workplaces, schools, commercial etc.

3. **User Classification**: User classification is similar to zone classification. Classification of user is based on their behavior. Users are classified into residents, workers, students, travelers, and consumers. One individual can be a resident in the night and traveler while going to work, and a worker while in office.

The key innovation of the paper is that the proposed model is flexible with respect to parameterization of user behavior while maintaining reasonable complexity.The model introduces generic classification locations and user behavior which allow derivation of the aggregated effects of user mobility. The model couples the user mobility model with traffic predictions to form a workload model.

Modeling of the zone was done using the zoning information found from the city development plans. Publicly available census data and public time budget studies were used to model the user classification and estimation of number of users in different zones. Estimation of people moving in and out of a zone and the number of travelers was obtained from the transport department.

Using the above information the authors were able to predict the number of users 'N' in a zone 'z' possessing a behavior 'b' at a time 't'. 'N' was multiplied with a time, behavior and zone dependent split factor 'f' to estimate the number of active users (users emitting traffic). Using the data obtained, the authors were able to estimate the traffic requirement in a zone at a time 't' or total requirements for a set of zones at any a time 't' or aggregated requirement over a day in a zone z.

The authors present a novel approach for developing a mobility model based on statistical data collection covering the macroscopic effects observed within metropolitan areas in context of heterogeneous wireless radio access. The model can be used to serve as a comparative analysis of quality of services routing strategies in wireless metropolitan area networks.

The model separates the influence of mobility and traffic to allow for greater flexibility. The mobility part is based on the combination of statistical zoning information with field data of movement patterns. This allows predicting the density of users classified into different groups for a given area at a given time. The combination of user density with the predicted synthetic traffic of the modeled user groups gives the traffic and fluctuations of traffic throughout the network, thus describing the workload for the envisioned scenario.

Few disadvantages of this approach are: A lot of coarse grained data and complex data mining is required for modeling. The statistics derived from a city may not be applicable in other regions or different countries because of different user behavior patterns and culture. The data may not be publicly available at all.

## 2.3 Adaptive Learning

*What limit we can put to this power, acting during long ages and rigidly scrutinizing the whole constitution, structure and habits of each creature-favouring the good and rejecting the bad ? I see no limit to this power in slowly and beautifully adapting each form to the most complex relations of life.*

- *Charles Darwin,* ***Origin of Species****.*

Learning is important for practical applications of artificial intelligence. Herbet Simon [3] defines learning as any change in a system that allows it to perform better the second time on repetition of the same task or another task drawn from the same population. The ultimate goal of adaptive learning is to develop computer systems that provide or support effective learning experiences for a wide range of learners across a wide spectrum of knowledge domain.

Today people interact directly, one-on-one, with their computers. As we move towards hundreds and thousands of computers per person, this paradigm cannot scale; placing too high a burden on individuals. With the proactive computing model, computers will anticipate our needs and sometimes take action on our behalf. We will continue to interact with a few of our computers but the vast majority will be embedded deep within our physical environment where they will capture and may act on data without human intervention. *David Tennenhouse* suggests that these future systems will have to be self-learning.

The need for the adaptive or self learning algorithms arises from the fact that the information received by a system may be unreliable or uncertain. Rules governing real-world events can never include all the factors that might determine whether their conclusions really apply. For example the correctness of basing a diagnosis on a lab test depends whether there were conditions that might have caused a false positive, on the test being done correctly, on the results being associated with the right patient [4]. In order to draw useful conclusions, future systems must be able to reason about the probability of events, given their current knowledge. The ability to learn must be a part of any system that claims to possess general intelligence. Intelligent systems must be able to change though the course of their interaction with the world as well as thought the experience of their own internal states and processes.

There are two aspects to the learning processes: Statistical Learning Theory and Machine Learning. The goal of Machine Learning is to automate the learning process and the goal of Statistical Learning Theory is to formalize it.

- **Machine Learning**

  Learning involves three-step inductive inference process:

1. Observe a phenomenon (sampling).

   (a) Construct a model of the phenomenon

(b) Make predictions

The observed phenomenon is analyzed for regularities in it. Generalization is made based on observed data and past knowledge. The regularities are then generalized from the observed past to the future. Predictions are made based on the generalizations derived.

- **Statistical Learning Theory**

  The fundamentals of adaptive machine learning are based on statistical learning theory. Statistical learning theory studies the process of inferring regularities from empirical data. The critical issue is defining generalization, i.e. what is called generalization: how it is possible to infer a law which will be valid for an infinite number of future observations, given only a finite amount of data? This problem relies on fundamental issues of statistics and science in general, such as the problems of complexity of explanations, a priori knowledge, and representation of data.

As mentioned earlier, information received by a system may be unreliable or uncertain. A very important theory to answer questions pertaining to partial and uncertain information about a system was proposed by Thomas Bayes. His theories on probability offers a mathematical foundation for reasoning under uncertain conditions and has become a major part of the mathematical foundations of application development. Bayes theorem is based on conditional probabilities; of an event is the probability of an event given some evidence. [5]

$$P(event|evidence) = \frac{P(event \, \hat{} \, evidence}{P(evidence)}$$

Although Bayesian probability the theory offer a mathematical foundation for reasoning under uncertain conditions, the complexity encountered in applying it to real life problems can be prohibitive considering the large number of events and evidences can come up in a real life problem.

Bayesian belief networks proposed by Pearl [6], offers computational model for reasoning to best explanation for a set of data in the context of a problem domain. Despite the name, Bayesian networks do not necessarily imply a commitment to Bayesian statistics. Rather, they are so called because they use Bayes rule for probabilistic inference.

## 2.4 Summary

This chapter described and analyzed various aspects or sensor nodes, routing protocols, simulation environments, design, implementation and performance constraints of sensor networks. We mentioned the related works that have influenced our approach towards solving the cane toad monitoring problem

Power constraints are of utmost importance and form the dominant research focus that is percolating through hardware design and network protocols. The sensor nodes for cane toad monitoring should be small and inconspicuous and not disturb the natural environment. To be effective as an unobtrusive cane toad monitoring tool, the nodes need to be ultra-small and ensure data delivery by dense deployment.

In the end we gave a brief introduction Learning theory, Bayes Theory and Bayesian Network.

# 3 Cane Toad Monitoring

## 3.1 History

There have been a many species in Australia that are not native to Australia but were introduced deliberately. Few have created a national ecological problem.

To name a few; cane toads introduced from Hawaii, rabbits from Europe, myna from India, European red fox, are species that are not native to Australia and are current or potential future problems.

The myna a bird was introduced in Melbourne in from India around 1860 has been found spreading fast. It has potential to become a big problem. Wild rabbits were introduced to promote rabbit shooting as field of sport. 24 rabbits were introduced and with in 100 years they multiplied into millions causing wide spread devastation.

Supportive climatic conditions, ample food, favorable breeding grounds and lack of natural predators were major contributors that influenced the uncontrolled growth of these species.

### 3.1.1 Introduction of Cane Toads

Australian farmers were troubled by cane beetles that were destroying the sugar cane farms.

Cane toads were being successfully used in Hawaii and the Caribbean islands to curb the growth of beetle, pests of sugar cane crops.

To stem the growth of cane beetles 102 cane toads were shipped from Hawaii to Australian in 1935. After being help in captive for a while the toads were released into the sugar cane fields.

Much to the dismay of the farmers it was found that the toads though very effective in controlling the beetles in Hawaii were not suitable for the cane beetles in Australia. The toads could only jump as high as 30 cm, which

Figure 9: Female Cane Toad

was not sufficient to reach the cane beetles who were perched on the upper stalks of sugar cane crop. Also, when the time of year came that served as breeding season for the beetles and their larvae was profusely found crawling from the ground, cane toads were nowhere to be found. The farmers resorted back to the use of chemicals to kill the beetles.

There were only 102 toads, no one cared to catch and dispose them. The impact on the beetles never resulted in a significant improvement for the sugar cane industry's crop loss, but did instead allow the cane toad species to multiply without limitations and greatly impacted the ecology of the area. The cane toads have innate ability to survive in adverse climatic conditions and variety of habitats as a result they multiplied at rapid rate overwhelming other species.

## 3.2 Cane Toads

Cane Toad is known as Bufo Marinus in the scientific world. The species classification is as follows: Kingdom Animalia, Phylum Chordata, Class Lissamphibia, Order Anura, Family Bufonidae, Genus Bufo, and Species Bufo marinus [45].

Toads of the family Bufo are found in Southern American countries, Australia, Africa and Eurasia. The cane toad is one of the largest toads in the world. Figure 9,10,12,11 show female cane toad, male cane toad, front view of cane toad and side view of cane toad.

Figure 10: Male Cane Toad

### 3.2.1 Cane Toad Identification

Cane toads are large heavily bodied amphibians. Females have smooth warty skin and males have rough skin like a sandpaper. Males and females possess football shaped pupil edged in yellow and hard bony ridges that meet above the nose. Males emit a distress call similar to that of a telephone dial tone when handled. Their mating call is a long loud purring trill. Male Cane Toads are smaller and wartier than females. The warts exist on both sides of the mid line of their back and end in dark brown caps.

The Cane Toads sit upright and move in short rapid hops similar to a frog.

- **Appearance**

  Their hind feet have leathery webbing between the toes and their front feet are unwebbed. All the toads have same appearance as that of a female until they become sexually mature and then males switch to the male type.

- **Color**

  They come in various shades of grey, brown and dark yellow depending on their habitat.

- **Size**

  The cane toad is one of the largest species of toad in the world, and can reach a possible maximum length of 238 millimeters, which is the

Figure 11: Side View of Cane Toad

highest recorded [16]. Average-sized adults are 10-15 cm long and weigh around 1.5 kilo grams.

- **Active Times**

  Adult toads are nocturnal, the tadpoles of cane toads have been found to be diurnal. Inactivity during the day prevents the loss of water from their body. During the hot and dry seasons

- **Preferred Habitats**

  If given a choice, they prefer to inhabit forested areas with water nearby, and therefore have plentiful insects as food. They are much more active at night, but can also be frequently seen hopping along sidewalks and resting in or near canals or pools of water. During the daytime hours and even in cold or dry seasons the toad will remain inactive underground in small shallow excavations or hide under debris, fallen trees, stones, or any other covering found typically outdoors [45].

- **Venom**

  When under attach from a predator the cane toad's skin releases a milky white venom.

#### 3.2.2 Reproduction

Cane toads have a typical amphibian life cycle involving the following stages: egg-tadpole-young-adult-egg. The number of eggs laid vary with the age of

Figure 12: Front View of the Cane Toad

the female. Mating can occur at any time of the year. Male toads attempt to mate with anything resembling a female toad living or dead.

Groups of male cane toads wait by the water bodies in the night and emit the mating sound, which is long sound of a phone ringing. The female toads loaded with eggs, travel towards the toads. She then chooses the male she wishes to copulate with. The copulation can continue at intervals for several hours or even days until the female is ready to drop her eggs.

After the mating, the male toad holds the female and they swim this way until the female approaches some type of water vegetation. She lays eggs in vegetation to conceal the eggs from predators. The female toad lays a long string of eggs coated in a substance similar to gelatin, while the male showers sperm over them. After the eggs are deposited, the female swims around the vegetated are in order to wrap the eggs around it.

The females have been found to lay 4,000 to 35,000 eggs at a time and may lay eggs twice a year. It takes 24-72 hours for the cycle from eggs to the tadpole stage depending upon the water temperature. The newly born tadpoles survive on algae and other aquatic plants until they mature into a juvenile.

The transition from tadpole to juvenile make take from three to twenty weeks after which they leave the water body and move on to the land. Temperature range of 25-30 degree C is ideal for healthy development. The juvenile takes around 1 year to become sexually mature.

Cane toad tadpoles are aggressive eater of the eggs and may consume 99% of the eggs laid. An extremely interesting aspect of Bufo marinus reproduction is that males are in fact able to reproduce without the presence of any female toads, because they possess a structure called a rudimentary ovary, in case their testes suffer damage or are removed. [16]

### 3.2.3 Comparison with natural Australian frogs

The toads are bigger than most of the Australian frogs, in fact they are one of the biggest toad species. The tadpoles of cane toads develop faster than those of the Australian frogs so they out compete the Australian frogs for food. Toads and cane toad tadpoles seem to be resistant to some herbicides and eutrophic water, which would normally kill frogs and frog tadpoles [16]. Most of the stages in cane toad's life are poisonous and therefore they do not have natural Australian predators to keep their numbers in check. The juvenile toads take some time to develop the poison glands and are susceptible in that stage. Australian frogs do not have any such defense mechanism in place. Cane toads not only eat the food normally eaten by frogs but also dog food, mice. They are therefore more likely to survive in adverse conditions.

### 3.2.4 Cane Toad Venom

One of the most important factors behind the high growth rate of cane toads is that they secrete a milky poison and are therefore poisonous to eat. This behavior is present at each and every stage of it's life cycle.

The cane toad venom is a complex mixture of 14 chemicals, which act on the heart and nervous system causing salivation, cardiac arrhythmia, high blood pressure and death. The source of the poison are the parotid glands and are perhaps the most characteristic feature of the adult cane toad. These glands are large, and are situated on the head, behind each ear.

A cane toad under threat turns it's side on to its attacker so that the venom glands face them. These glands not only ooze venom but also can throw it over a distance of 2 m if the toad feels it is in high danger. Cane toad venom can be found all over their skin. Predators picking up a cane toad and receiving a dose of venom may die within fifteen minutes.

Humans are not greatly affected by the cane toad venom but if this venom reaches the eyes it can cause irritation or temporary blindness.

### 3.2.5 Dietary Habits

Cane toads are omnivorous; they will prey on a large variety of animals and eat anything smaller than itself. Cane toads under captivity have been observed to eat everything from dog food to mice.

Although they are primarily insect feeders, they will attack anything that moves and is small enough to fit in their mouths. They have been known to consume small vertebrates, mollusks, arthropods, plant matter. Their diet includes insects, termites, beetles, small lizards, frogs, mice and even younger Cane Toads. The cane toads both hunt and lie in wait for their prey to come in the range. Once the prey is with in the range the toads catch the prey with their sticky tongue and swallow it whole.

Cane toads have also learnt to eat domestic pet food. They are found to have stolen food from dog and cat bowls. They have been found to eat thrown meat, discarded vegetable matter, and human feces.

### 3.2.6 Habitation Environments

Cane toads have been found to survive in a variety of habitats. An analysis of the ecosystem of Kakadu and the variety of habitat where cane toads have been observed to survive, suggest that 90% of Kakadu is fit to be habitated by the Cane toads.

### 3.2.7 Life Cycle

In this section we list the habitats for the various stages in the life cycle of cane toads.

- **Embryos and Tadpoles**

  Embryos have successfully developed into tadpoles in range of habitats in which include sea water, temporary and permanent water bodies(for both slow and running water). Preferred habitats consists of shallow transparent water bodies with abundant sunshine, this is attributed to presence of algae in such habitats. As tadpole matures into a juvenile toad it's swimming ability decreases and therefore it prefers shallow water.

- **Juvenile Toads**

  Metamorphosis of tadpoles into a juvenile is much faster for cane toads than the native Australian frogs. Their small size makes them vulnerable to water loss, as a result the juvenile toads are active close to water bodies and moist regions. Preferred regions are regions with sufficient moisture, vegetation cover for safety and food resources. As the toad matures into an adult, it's size increases and hence the ability to carry and retain water. The ability to carry water allows the adult toad to move away from water bodies and it becomes more nocturnal.

- **Adult Cane Toads**

  The Cane toads original habitat before its dispersal by humans was subtropical forests near fresh water. However, they have adapted and they shelter in ponds, moist crevices and hollows, excavated depressions , beneath logs, rocks, debris, junk, gardens, drainage pipes, cement piles. They can also survive in saline water bodies though it is not a preferred habitat for them.

  The adult toads can survive the loss of up to 50% of their body water, this adds to the options of shelters they can take refuge in.

Cane toads will usually stay on dry land; and reproduce in any shallow water near its surrounding. They prefer areas disturbed by people and farm animals. Flattened areas make it easier to get around. Large toads can be active in cool, dry shady or windy night time conditions that inhibit activity of small cane toads. Adult cane toads are mostly nocturnal during the warm months of the year. In the dry periods they keep their shelters moist, sometimes through urination.

#### 3.2.8 Behavior

Toads huddle together in groups to stay damp during the day. At night they become active, mainly eating and breeding and are relatively solitary.

### 3.3 Impact

Even though the cane toads have failed in its original purpose, it is succeeded in becoming one of the Australia's leading pests. It has also been successful in becoming an ecological threat to the environment and it's competent species.

#### 3.3.1 Impact on Animal Kingdom

Despite the great body of knowledge that exists on the general biology and ecology of the cane toad in Australia, there have been a few quantitative studies to investigate it's impact on Australian native vertebrate and invertebrate fauna. In contrast there is substantial anecdotal and experimental evidence to suggest that cane toad do impact adversely on the native fauna. They have been successful in becoming an ecological threat to the environment, because of their high competence.

In this section we discuss the impact of cane toads on the it's prey's, competitors, predators and a few ecological impacts.

- **Animal Kingdom**

The cane toad's venom is poisonous to other animals. The parotid glands of the toad release toxin when the animal is provoked or squeezed, such as when in the mouth of a predator. Experts believe that they already have decimated water rat, quoll, native frog and some bird populations in Queensland and are capable of wreaking an ecological disaster in the fragile Kakadu ecosystem [15].

- **Native Frogs**

  In the wild, Australian native frogs and cane toads and compete for food, shelter and breeding sites. The cane toads have higher growth rate than the tradition Australian frogs and thus outnumbering and out competing them for food and shelters. This is said to have caused numbers of some native frogs to drop.

- **Predators**

  Many native Australian frogs produce mild toxins that aid in defense, but the potency of cane toad toxic makes it lethal to most eating animals. The published research work in this direction is less or not publicly available. The possible predators at risk consists of fishes, lizards, snakes, crocodiles, turtles, birds, quoll (a native cat).

- **Prey**

  The adult cane toads try to eat any thing which is smaller than itself and consume a wide variety of prey. Major prey include beetles, ants, termites.

### 3.3.2 Other Impacts

- **Contamination of Water**

  Larvae and dead cane toads are often found dead in water bodies, with their skin covered with the milky white venom. This will cause the contamination of drinking water for the aboriginals. Snails have also been found to die in water having cane toad toxin.

- **Botanical Effects**

  Though not analyzed in depth, the role played by ants and termites in the Kakadu ecosystem will be affected. These roles are associated with seed harvesting.

- **Human Health**

  The cane toads have been found to eat human feces, as a result they carry bacteria that cause gastroenteritis, typhoid and septicemia. The presence of cane toads in large number can be a health hazard [16].

Cane toads are advancing progressively in an unsuspecting Kakadu environment, it is predicted that they will come close to wiping out native quoll, threaten the existence of few native frog species, and consume the food supply of many other animals. There appear to be few predators for all life stages of toads, mainly because of their poison glands. They will thus vastly affect all the animals below, above or at the same level to them in the ecological tree.

## 3.4 Cane Toad Monitoring Importance

The toads have little if any predators in Australia to control their population and have therefore multiplied in densities ten times of those found in native habitats. One of the reasons Cane toads is so successful in varying environments and when introduced to new areas is that they possess a remarkable ability to adapt to a wide range of habitats. [16].

One study conducted in 1990 by the Commonwealth Scientific and Industrial Research Organization showed that these toads are very effective competitors with other insect-eating animals, were highly toxic to many possible potential predators, and also could have a negative impact on other native frog species.

Kakadu is rich habitat for the cane toads. It has ample water resources and due to the low soil fertility of the lowlands of Kakadu, insects are preeminent amongst animals, and therefore food is in plenty. To counter the growth of cane toads the Australian government granted the Commonwealth researchers one million dollars from the National Heritage trust so that research could be initiated to see if modern gene technology could be used to somehow deter the toads from marching all over the country. [16].

### 3.5 State-of-the-Art Research

- Impact of Cane toads on native frog species

  Andrew Taylor from the University of New South Wales has developed a vocalization recognition software to census native frog communities. This will provide data related to impact of Cane toads in the native frog communities. This is made possible by the plebs that record rainfall, temperature and distinctive vocalizations or Frog Calls during the wet seasons when the frogs are particularly active. The pleb consists of 25MHz Intel 486 CPU, flash memory to store data, solar panel for energy, microphone . The pleb is able to record the presence of around 22 frog species present in the area. To determine whether the observed vocalization are of the specified type, the system first produces a spectrogram using Discrete Fourier Transform. Every species has it's own characteristic spectrogram which is known to the system. The observed spectrogram is compared with the known spectrogram in the system, and a decision is made weather the observed vocalization belongs to any of the species known to the system.

  [17] Calls are detected in the spectrogram using an ad-hoc combination of local peak finding and a connectedness measure. Attributes are extracted both globally from the call and from a window moved incrementally through the call. Quinlan's C4.5 machine learning system is used to induce a decision tree-based classifier monitoring applications.

  This software system will be used in unattended operation to monitor the effect on frog populations of the introduced Cane Toad. This how ever has its drawback, one such pleb costs 1000 Australian dollars and it will be very expensive to deploy them all around the Kakadu national park. Even though it collects the data automatically some one has to go to the pleb to retrieve the data. Information is not real time and is only available when some goes and collect it. It can be used to detect the presence of cane toads but wont be able to infer the direction of movement of cane toads.

- CSIRO (A gene that will stop cane-toad growth)

At Commonwealth Scientific and Industrial Research Organization (CSIRO) researcher are planning to put a virus into the gene which will either kill the tadpole or make the cane toad infertile. The scientists have to ensure the gene can be transmitted without affecting any other species.

Earlier, Venezuelan viruses were tested on the cane toads and were proved effective in killing cane toad tadpoles; they also killed one species of Australian frog in the trial and were therefore rejected.

Recent work has been to develop a gene, which if introduced into the tadpole will stop the growth of tadpole into an adult cane toad and thus preventing any further breeding. The gene consists of large, double-stranded DNA viruses. They are very hearty viruses that are well equipped for infection, and may readily be able to infect cane toad tadpoles.

Because the biological differences between the tadpoles and the adult cane toads, there is a good chance of identifying the gene that is critical for the toad's metamorphosis from tadpole to adult. Researchers believe that the toad's immune system will see the gene product as an unwanted invader and will initiate an immune response against it. This response should inactivate the gene product, preventing the toad's transition to adulthood. All this is possible because the working of immune, digestive and blood systems are vastly different in tadpoles compared to cane toads.

## 3.6 Long-term Research Goals

Traditional monitoring techniques require lot of human intervention, which is not highly desirable. Few species might react adversely to human presence. A lot of the species are hard to locate and are detected by their vocalizations. Humans might not be present when animal calls are made and when detected the calls are prone to error in observation.

Other habitat monitoring technique is to use one or a few sophisticated instruments called plebs [46]. These plebs are are typically large in size, expensive and do not provide real time information. Due to the large size and costs, the plebs will be sparsely deployed. This would create gaps in the information received requiring the biologists to make generalizations.

Long term research in this field should able to do the following.

- Track the direction of migration of Cane Toads.
- Growth of movement of Cane Toads in a region and over all growth in the park.
- Should be able to infer the impact of cane toads invasion on the flora and fauna.
- Vocalization techniques should be used to measure change in numbers of predators, preys and competitors.
- Pinpoint the regions habitated by cane toads. This kind of information will be needed by the biologist to selectively inject the toads with the virus. This can be done using GPS or localization techniques.
- Development of a system which works with out human intervention, be robust, scalable and have a long battery life.
- The instrument deployed should be inconspicuous, not disrupting the area under observation.

## 3.7 Summary

If allowed to grow freely, Cane Toads will cause havoc in the Kakadu national Park. They have the potential to colonize the whole of Kakadu National Park and may cause extinction of few species in that region.

Kakadu national park is a rich habitat with ample food resources, plenty of breeding sites for the Cane Toads. The floodplain, swamps, monsoon forest, riparian, woodlands, open forest, springs, soaks and waters pools will be the preferred habitat of Cane Toads in the Kakadu national park. A wide variety of species in the animal kingdom will be affected. There will also be species, which will be indirectly affected. Northern quoll, snakes, native frogs, lizards will be affected and need to be monitored as well. There is one advantage; because of the cane toad invasion the population of feral cats is expected to decrease.

Although there is still much work to be done in areas where cane toad continues to spread, there are reasons for optimism as far as the control of the problem. In areas where the toads have been around for the longest times known to this day, their population numbers tend to decline after the initial breeding sprout and increase. One of the reasons is that some native animals are learning to get around the toxins and have adapted to eating them.

There are important research gaps pertaining to impact of Cane Toads that need to be filled. It is hard to estimate the impact of Cane toads on the ecosystem because of lack of metrics and models available. A dense deployment of sensors in the Kakadu National Park as discussed in following section can provide coarse grained information that was previously impossible to obtain using traditional methods.

# 4 Large-Scale Sensor Deployment

The objective of this Habitat Monitoring case study is to conceptualize the deployment of an array of inexpensive, lightweight sensors that are capable of acoustical observations to monitor, track and characterize the impact of Cane Toads in Kakadu National Park. They have the capability to sense and interpret information, log data and later transmit core data to a central facility. The central facility receives this report which contains several environmental parameters across a range of temporal ecological gradients.

The ecologist or the biologist rely on generalizations and extrapolations of measurements at sites or other similar monitoring studies removed from the actual location of study. This system will be able to give detailed information with minimum delay to the researchers. Finally this system will operate continuously in remote areas in Kakadu National Park for a number of months or years without any support, and be capable of delivering data for years. The kind of support envisaged is replacement of batteries on which these sensors will run.

## 4.1 The Problem

As mentioned in the previous chapter, cane toads will be invading Kakadu National Park, one of Australia's most prized natural environments. Kakadu will be a rich habitat for Cane Toads and therefore it is imperative that their activities are properly monitored and recorded.

We hope to equip biologists and ecologists with a tool to track and monitor the growth and impact of cane toad populations. The Australian government currently periodically release viruses into the environment to control the growth of the rabbits. The viruses are released from a helicopter irrespective of the fact that rabbits may not be present in that region, sensor networks can bring such costs down. Sensor networks will come in handy if selective cleansing of cane toads is required within Kakadu national park. The fully deployed system will be capable of pin pointing the count and concentration of cane toads in the Kakadu national park.

## 4.2 Design Goals

The basic design goal for this study is to provide cost efficient sensor coverage of Kakadu National Park while guaranteeing a certain Quality of Service.

Kakadu National Park is huge. Deployment depends on a lot of factors like energy, data rate, processing power, active periods, idle periods, distance between nodes and range of sensing. The approach for deployment will be to deploy it in a part of the park, learn from it, improve the design and deploy it in the next region. The process is iterative.

In the following two sections we will discuss various functional and performance characteristics expected from the system.

### 4.2.1 Performance Characteristics:

1. **Scalable**:

   The network should be scalable from a region of park to have sensor coverage of a wide area of the Kakadu National Park which is 20,000 sq km.

2. **Energy Efficient**:

   Longevity of a sensor network is dependent on how energy efficient it is. To get the maximum of the network the system needs to be energy efficient.

3. **Robust**: The system should be tolerable towards

   - Networking errors
   - Data Transmission Errors
   - Sensing Errors

4. **Fast data processing**:

   The system should be able to process data quickly. The network load is expected to be quite high during peak times and close to zero rest of the time. Grigg et. al. performed similar experiments and it was found that 15 seconds of sound takes 1 minute of process.

5. **Optimization**:

   The system should programmed to optimize the trade off between data processing and transmission. An important question would be whether data should be processed at micronode or macronode or a combination of both.

6. **Adaptable:**

   Ultimately, the success of the network depends on how adaptable it is. How much it learns from past mistakes and improves.

### 4.2.2 Functional Characteristics:

1. **Storage for data**: The cane toads are most active in the night and early part of the day, it makes sense to keep storing the data in the night and then transmit the aggregated data in the day.

2. **Detection Accuracy**: The it impossible for the system to learn from cane toad vocalisations under laboratory conditions and give 100% positive detection when deployed. Under deployment conditions the vocalizations from the cane toad will be corrupted by noise from frogs, rain, wind and other animal calls.

3. **Target Localization:** The system should be able to point the location of the cane toads with a certain degree of accuracy.

## 4.3 Domain Knowledge

Statistical information which would help us predict the presence of cane toads is not available. We lack data on the presence of mud holes, creeks, temporary water bodies, disturbed regions etc. Therefore, we have used very rough coarse grained metrics to select zones fit for sensor deployment.

The following set of data, was the basis behind several of our assumptions and decisions.

a) Within an occupied habitat, cane toads spread at a rate of 100 km/year.

b) The overall spread of cane toads between catchments is 27 km/year. i.e. between habitats.

c) The toads need to visit water once in three days.

d) The toads are most active from 9 PM to 1 AM in the night, and from 3 AM to 5 AM in the morning.

e) They cluster together in groups in the day time to stay moist and are likely to concentrate near the water bodies.

f) Cane toads can disperse by natural and by transport mediums used by humans. Roads and vehicle tracks have been known to provide routes to toads for traversal. Such regions also have open level ground along which the toads concentrate their activity. The major transport means for cane toads will be "hitch-hiking", in vehicles, caravans, or removal containers and natural dispersal along transport corridors, most likely the Arnhem and Kakadu highways and major rivers and creeks [16]. Faster natural means include transport of eggs and tadpoles by flowing water, swimming by adults in flood water and hiking.

## 4.4 Solution Approach

### 4.4.1 Assumptions

- **Homogeneous nodes**:
  - All micronodes(sensor nodes) are identical in resources and capabilities and similarly all macronodes(plebs) are assumed to be identical.
  - Static environment: the environment is assumed to be static, at least to the extent that gross topology remains unchanged.
- **Model**:

  There are no prior models or research work available for this type of environment. This algorithm is intended for applications in which environment models are unavailable, incomplete or inaccurate. A key task for the network will be to generate a model.
- **Localization**:

  The position of each and every node is known in some arbitrary global coordinate system using any of the proposed localization techniques.

The proposed solution consists of 4 steps.

1. **Zone Division**: Division of deployment area into zones.

2. **Zone Categorization:** Classification of zones based on deployment priorities.

3. **In-Zone Deployment**: Strategies for deploying nodes within a zone to meet the bandwidth and coverage requirements.

4. **Adaptive Learning Algorithm**: It is hard to get initial deployment right due to uncertainty. Bayesian framework is used for handling uncertainty in domain knowledge and using it to drive adaptive learning algorithm

We will discuss each one of them in the following sections.

## 4.5 Zone Division

The area of Kakadu National Park is around 20,000 sq km. We divided it into 2000 regions of size 10 sq km each. The basis behind 10 sq. km is that we feel that is small enough to experiment and learn from the first deployment of sensor network and large enough for the macro effects (like effect of water bodies and food resources on cane toad population) of the application to influence the presence of cane toads in the zone. These zones are categorized based on which are most likely to be "hotspots" for cane toads gathering as discussed in the following section.

## 4.6 Zone Categorization

The whole of Kakadu National Park was broadly divided into zones. Zones were categorised into three based on which are most likely to be hotspots for cane toads gathering.

a. Highly Probable.
b. Probable.
c. Not Probable.

An important step would be to categorize these zones further during the iterative learning stage(Section 4.8.1 ) which comes after the 1st deployment. Due to the lack of metrics available to further categorize these zones, categorizations has been made broad.

An analysis of all the 2000 zones showed that all of them were fit to be colonized by the cane toads. The zones were divided into Highly probable and Probable based on the availability of water bodies in the zone. None of the zones was put into the Not Probable category.

## 4.7 In-Zone Deployment Strategy

This section provides a description of the wireless sensor network deployment for real time cane toad monitoring. The system architecture addresses the performance and functional requirements for habitat monitoring in Kakadu National Park.

### 4.7.1 Overview

The two requirements of our habitat monitoring system are target detection and target localization. The first is to determine whether observed animal calls belong to cane toads using their spectrograms (species detection). The second is to locate the position of a cane toad when its vocalization is recognized (target localization).

Most of the cane toad vocalizations function as an advertisement to other members of the opposite sex and hence are species-specific. A variety of properties can be used by the system to recognize the vocalizations of cane toads. These include call rate, call duration, amplitude-time envelope, waveform periodicity, pulse-repetition rate, frequency modulation, frequency and spectral patterns [12].

Target location refers to pin pointing the location of the target two or three dimensional grid at any point in time. Target location identifications have a few problems. For instance, for enhanced coverage, a large number of sensors are deployed in the field and if the coverage of sensors area overlap, they may all report a target in their respective areas. It may be hard to pin point the exact location considering the granularity of the grid in the system. Target location can be simplified considerably if sensors are distributed in such a way that every point in the grid is covered by a unique set of sensors. However this is hard to achieve as such a large number of nodes can not be placed manually or with a robot with precision in a natural environment like ours. The Kakadu National Park is full of water bodies, trees, vegetation and cliffs.

We plan to use the approach similar to Wang et. al. [32]. They proposed preprocessing of sensed vocalization data in habitat monitoring applications. They have suggested a 2-tier network for the purpose of collaborative signal and information processing. Densely deploying a zone with only sensor nodes brings in a lot of compromises. The individual nodes have to do processing, communication all by themselves and transfer the data back to the base station. This puts a big energy load on all the nodes. We believe using a tiered platform consisting of a heterogeneous collection of micronode, macronode and base station will be a much more effective design. Larger, faster and more expensive macronodes can be used more effectively by also using smaller, cheaper, and more limited nodes (micronodes). Since a lot of heavy weight processing is done at the macronodes, clustering will drastically relax the resource requirements at the base stations. Figure 13 shows a sensor patch with data transfer along the micronodes towards the macronode.

### 4.7.2 Components

All nodes have integrated capabilities of sensing, processing, and communication. However, real-time processing is a big challenge for resource-constrained sensor nodes. Acoustic signals are sampled at a rate of several KHz. This makes the censusing of animals which make frequent distinctive vocalisations expensive and time consuming. The toads are active in the night and they tend to repeat their vocalisations incessantly and may call in choruses with tens of toads present. The nodes will have to cater to more than one vocalization at a given point of time. Therefore, it is too demanding and time-consuming to conduct target classification and localization whenever a new sample is obtained.

We propose to use two types of nodes. Micronodes at the lowest level and macronodes at the higher level.

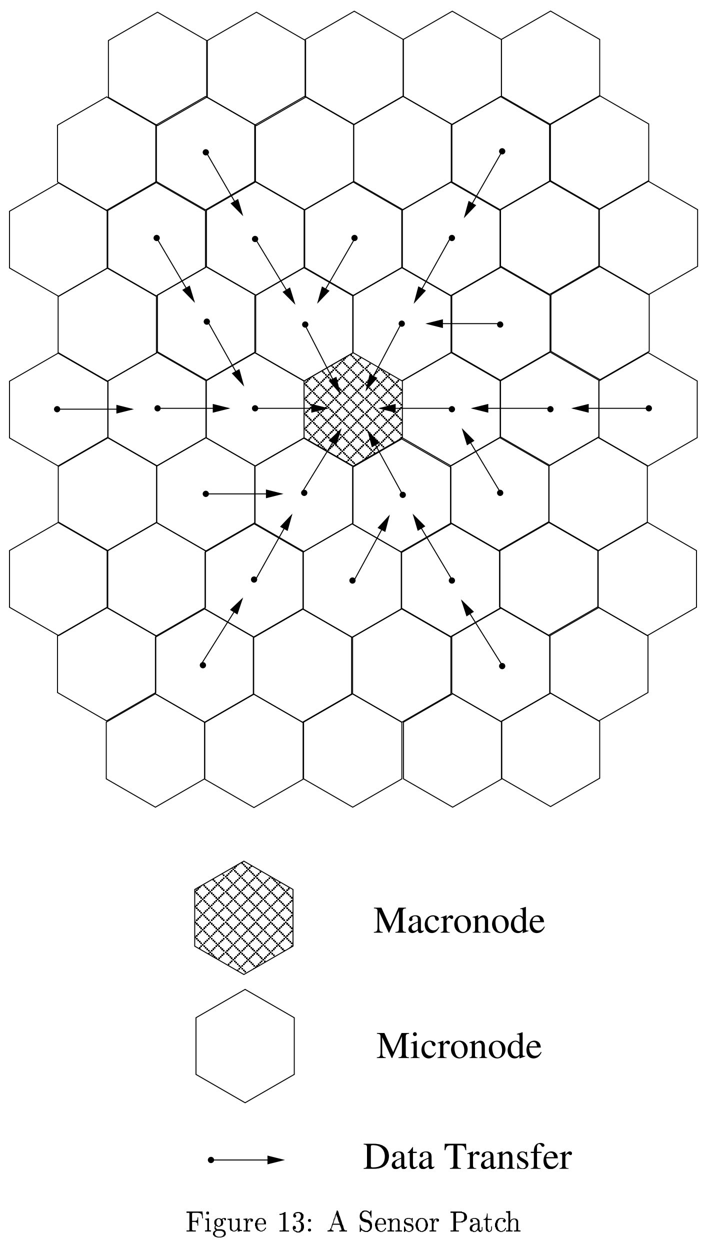


Figure 13: A Sensor Patch

Micronodes are less expensive and more resource-constrained than macronodes. Micronodes will be deployed in large numbers, where as macronodes will be limited in quantity due to higher power requirements and cost constraints. For target localization the Macronodes will be GPS enabled. The use of GPS is not desirable for micronodes because of high power consumption, antenna size and overhead obstructions. Both micronodes and macronodes will have acoustic sensors and communication will be over a wireless network.

The deployed network will consist of the following:

1. Micro nodes (lightweight, power, processing, memory constrained sensor nodes)
2. Macronodes (Plebs with more power, processing, memory)
3. Base Station (unconstrained powerful machines, link between sensor networks and the wired network)

Micronodes will be densely distributed because of their low cost. High density of micronodes increases the probability to detect cane toads and target localization is more precise. Macronodes are sparsely distributed because of their cost. Base stations will be less sparse than macronodes. The Kakadu National Park is big, therefore we need to have more than one base station. Micronodes, macronodes and base stations will form a three tired clustered wireless network(Figure 14) .

#### 4.7.3 Deployment Strategy

The three tiered architecture has micronodes at the lowest level, macronodes at the middle and base stations at the top. Sensor nodes will typically be deployed in dense sensor patches that are widely separated. Each patch encloses a particular geographic region of the Kakadu National Park.

Individual sensor nodes communicate and coordinate with one another in the same geographic region. This coordination makes up the sensor patch. The sensor patches are typically small in size compared to the size of the

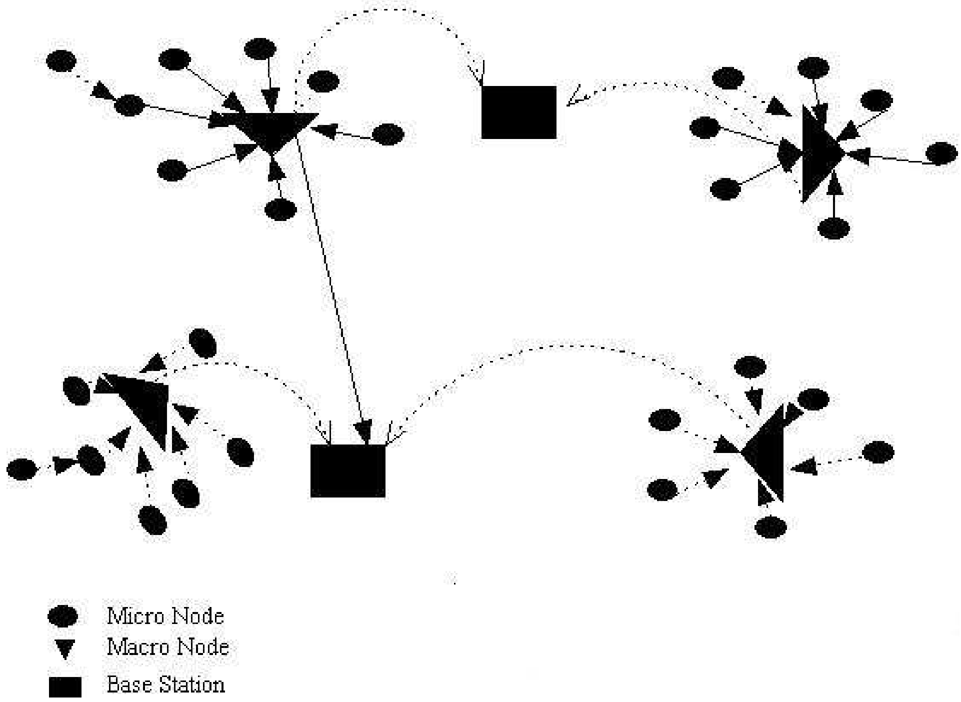


Figure 14: Sensor Network

park. We estimate the order of magnitude of the patches will be in thousands, and one zone can have one or more patches . The micronodes are responsible for sending the collected data to a centralized authority called the cluster head which is a macronode. One micronode can be only under one cluster head. The cluster heads are responsible for sending the data to the base stations.

When the system initializes the nodes are organized into clusters. Clustering techniques can greatly reduce the energy dissipated and can be achieved automatically by self-assembly [9]. The cluster head is used for collaboration and central data processing. All other nodes in the cluster are for distributed sensing and data preprocessing. Heinzelman et. al. [29] showed that clustering is most efficient for static networks where data is continuously transmitted. We expect this to be the case for our system, as cane toads have fixed active periods and incessant vocalizations are expected. Unlike LEACH, our cluster heads will not be rotated.

The number of macronodes or cluster heads will be limited, bounded by cost constraints and bandwidth requirements.

### 4.7.4 Sensor Node Dynamics

To extend the network, lifetime sensor nodes are expected to spend most of the day sleeping. The sensor node will periodically wake up to sample, compute and communicate in the night.

Cane toad species are most active from 9 PM to 1 AM in the night, and from 3 AM to 5 AM in the morning and utter a long loud purring trill and serves the purpose of mating advertisement. The probability of vocalizations is much higher at these times and least at other times.

The sensor nodes will be preprogrammed to coincide with the active and idle times of the cane toads. The nodes will have sleeping periods all during the day and will be programmed to wake up and start listening for vocalizations as the active time approaches. In the active times all the nodes may not be necessarily active. They will periodically go to sleep if no cane toad activity is detected.

Minimizing power in sleep mode involves turning off the sensors, the radio, and putting the processor into a deep sleep mode. I/O pins on the microcontroller need to be put in a pull-up state whenever possible, as they can contribute as much as 100 $\mu$A of leakage current.[22] A challenging issue will be to prevent leakage. Leakage dominates the energy resources required for nodes in sleeping mode. Several architectures like MICA use a DC booster to provide stable voltage from degrading alkaline batteries. With no load, the booster draws between 200 and 300 $\mu$A, depending on the battery voltage[22]. The nodes should reliably bypass the DC booster while reducing the supply voltage in sleep modes.

### 4.7.5 Spatial Density

Sensing function generally depends on distance of the source from the microsensor node. The specific sensing function parameters depend on the

| Detection Rage | Area Covered | Spatial Density |
|---|---|---|
| 20 m | 1040 sq m | .001 nodes/sq m |

Table 2: Spatial Density

nature of the sensor device and usually have the form $\mathrm{d}^{\mathrm{k}}$.

$$density \propto d^k$$

d = distance between the source and the microsensor

$k$ typically ranges from 1 to 4.

The sensing ability is defined as:

$$S(s,p) = \frac{\gamma}{d(s,p)^k} \tag{1}$$

$S(s,p)$ = the sensing ability.

$d(s,p)^k$ = the euclidean distance between the sensor node s and the cane toad at point p.

$\gamma$ and $k$ are the sensor technology dependent parameters.

Cane toad vocalizations can be detected successfully by the micronodes up to a distance of 20 m. Detection range of 20 m is the dominating factor in judging the distance between two nodes, as it is smaller than the transmission range of a typical sensor node. Table 2 gives the minimum node density and the area covered by one node (with $k = 2$) to be able to detect toads in a zone.

### 4.7.6 Vocalization Processing

The cane toad vocalizations are detected by the micronodes which are at the lowest level of the model. These micronodes perform preprocessing of the vocalization to determine if vocalization belongs to the cane toads. When a vocalization is received it is sampled and a spectrogram is created. The decision that the vocalization is of the species to be observed is determined by

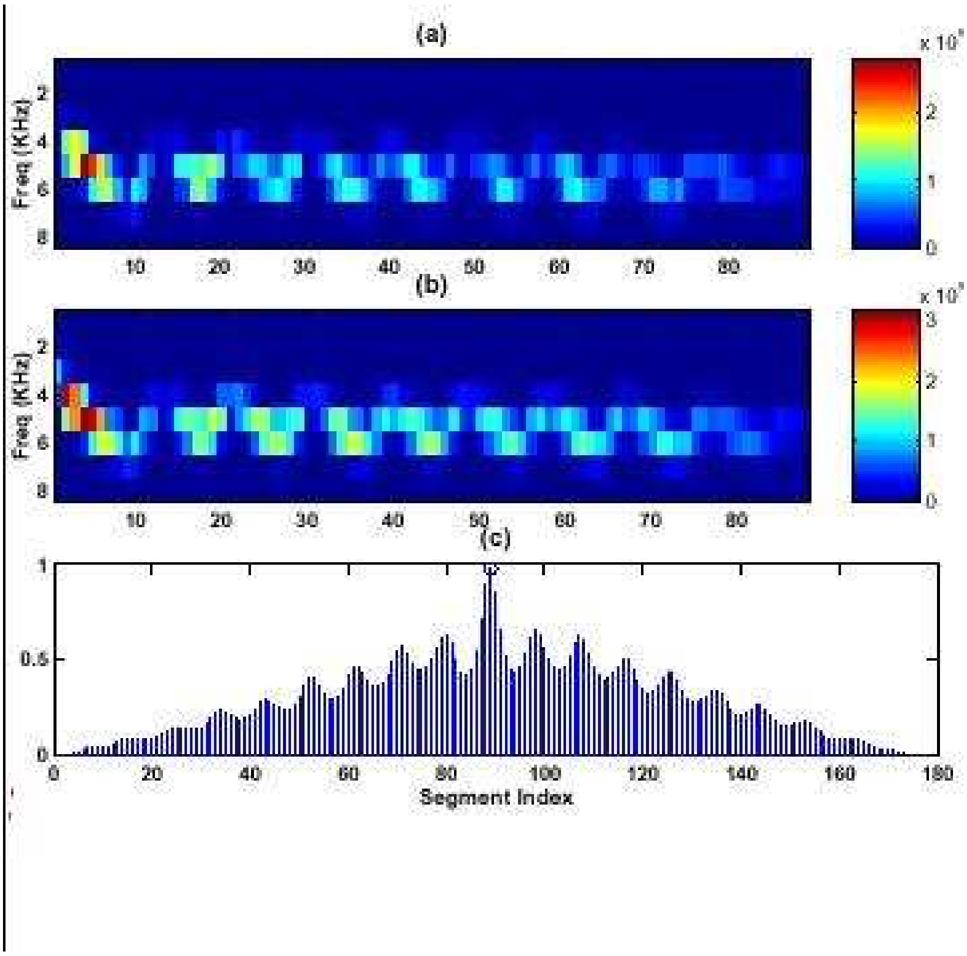


Figure 15: Spectrum of Frog

the maximum cross-correlation coefficient between the observed spectrogram and the specified characteristic spectrogram which is input to the system.

Figure 15 (a) shows the spectrum of observed vocalization, Figure 15 (b) shows the spectrum that is input to the system and Figure 15 (c) shows their cross-correlation coefficients. (source: [12])

Errors can be induced into the cane toad vocalization detection because of the background noise. Background noise can be in form of rain, wind (wood, bushes) and other animal and bird calls. Noise from wind/woods/brushes usually have a different frequency from that of animal calls because the evolution favors species that can make themselves heard clearly. The noise that

| **Frequency** | **Bits/Sample** | **Data Rate** |
|---|---|---|
| 10 KHz | 8 | 160 kbps |

Table 4: Bandwidth Requirements

matters is the one that in the same bandwidth as the animal calls. Table 4 shows the bandwidth requirements to process a cane toad vocalization

If nature of the noise expected is known, the system may add the noise at a certain SNR level to a clear animal call and then compute the similarity between the noise-corrupted call and the clear call. When the computed similarity is greater than the projected threshold, it is a detection otherwise, it is a miss.

We can use multi stage event-driven processing in order to reduce the processing and transmission cost . Acoustic data are processed in stages, starting with the least expensive stage. Min et. al. [40] show that if the preprocessing is done at the micronode and then the data is transmitted to the cluster head, it is less expensive in terms of total system energy required for processing. Acoustic data is processed in stages, starting with the least expensive process. Data is moved to the cluster head for complex resource intensive processing only when it passes the current stage of processing.

The whole data processing task is divided into three stages. From the fastest to the slowest: signal intensity monitoring, target classification and target localization. Signal intensity monitoring is fast and runs all the time on the cluster head. The micronodes continuously sample acoustic signals and buffer the last several seconds of vocalization. The observed spectrum is compared with the input spectrogram of cane toads only when the observed signal intensity exceeds the input threshold. If the vocalization is classified as the specified type, the cluster head estimates the target location using TDOA-based beamforming. Such staged event-driven processing will save time and energy because unnecessary processing of irrelevant acoustic events are avoided . The event driven processing approach was proposed by Wang et. al. [11].

### 4.7.7 Localization

Macronodes will have their locations, estimated using GPS receivers. They will act as the reference nodes for the micronodes under them. The location of the micronodes is determined using triangulation. A object in a two dimensional space can be uniquely positioned when atleast three reference points that form a are associated with it. After the macronodes estimate their positions, a distributed process of iterative triangulation starts, where each node that estimates its location becomes a reference point for other nodes. This process allows for a maximal number of nodes to approximate their location having only a minimal number of initial reference nodes.

To locate the calling animal when its call is recognized, techniques like finding the location through multiple angle of arrival measurements of beacon nodes [33] could be used. A sensor sends a beacon message and neighbors use trilateration based on signal strength measurements. The system proposed by Wang et. al. [12] determines the target location by Time Difference of Arrival(TDOA) based beamforming. Cross-correlation between waveforms of the same signal recorded by two different sensors indicates TDOA between those sensors. Given locations of multiple sensors and TDOA among them, the target location can be estimated using the least square method [13].

Beamforming is a technique that is be used for noise mapping from a distance. Beamforming uses a spatial filter that operates on the output of a set of sensors. The response of a given micronode is plotted on a polar graph, where the angle is the offset from the Maximum Response Angle(MRA) with the radius as the magnitude of the response (dB) in that direction. The results from multiple sensor nodes are used to achieve a narrower response in a desired direction (the MRA) [39]. When a cane toad call is recorded as a sound in a given beam, we can estimate which direction and distance it originated from.

The localization depends on the coverage of the region by the sensor nodes.

The sensing ability is defined in equation 1.For a sensor node $s_i$ it is given by:

$$S(s_i, p) = \frac{\gamma}{d(s_i, p)^k}$$

The Coverage for the whole zone is defined as:

$$\sum_{i=1}^{n} S(s_i, p) \tag{2}$$

### 4.7.8 Data Compression

Even though micronode processing avoids transmitting raw data to cluster heard by processing data locally, the beamforming nodes still need waveform data transmitted from multiple sensor nodes. Data reduction and compression techniques need to be used before waveform data are transmitted to a beamforming node. Data reduction lowers data volume by discarding irrelevant information in the waveform data [11] .

## 4.8 Adaptive Learning

It is impossible to get the deployment right the first time. Thus, it is prudent to have a self learning algorithm. We use Bayesian networks to learn from previous deployments and improve the next deployment. The objective behind a Bayesian network implementation is to bridge the gap created because of intuitive probabilities and lack of data .

### 4.8.1 Bayesian Belief Network

Murphy [31] in his lecture states "*A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest*". When used in conjunction with statistical techniques, the graphical model has several advantages for data analysis. One, because the model encodes dependencies among all variables, it readily handles situations where some data entries are missing. Two, a Bayesian network can be used to learn causal relationships, and hence can be used to gain understanding about a problem domain.

We use inferential statistics using a Bayesian Network to make valid predictions based on only a sample of all possible observations. The variables of a Bayesian belief network have been determined by exploiting the domain knowledge and self learning from obtained data. Domain knowledge has the benefit of the experience of the experts and often it is expensive to design a network without the domain knowledge or the probabilities set are inaccurate. Network topologies perform optimally when they are linked closely to the topology of the application and exploit the domain knowledge.

Learning from obtained data is challenging and has a few problems. Some statistical information is not readily available for learning purpose. In this case it is not even possible to create an environment of the type of Kakadu under laboratory conditions and generate data for learning. This lack of data in the design phase can cause serious miscalculations in the deployment phase.

A big challenge is to adapt the application to abrupt or slowly changing environmental characteristic with time, for example, seasonal influences like wet seasons when Toads are highly active and breed compared with dry seasons when they go into hiding to prevent the loss of water from the body. The principal reason behind using Bayesian Network is to bridge the gap between probabilistic models and human intuitive approaches to modeling uncertainty.

### 4.8.2 Conditional Probability

If A and B are two exclusive but possibly dependent random events, then the conditional probability that A is true, given B is true is defined as:

$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ if $A \cap B \neq \Phi$

$P(A \mid B) = P(A)$ if $A \cap B = \Phi$

$P(A \mid B)$ is defined as the fraction of the system in which A is true given that B is true.

The black region in the figure 16 represents the world where both A and B are true.

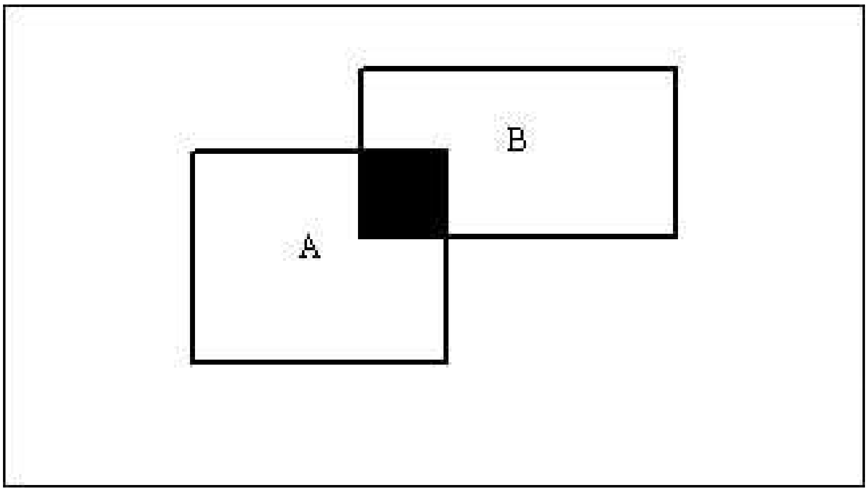


Figure 16: Region where A and B are true

Corollary:

$$P(A \cap B = P(A|B) \times P(B) \tag{3}$$

Similarly,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{4}$$

$$P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)} \tag{5}$$

The above equation is obtained by replacing the value of $P(A \cap B)$ from (3 ), is known as Bayes Rule.

**Total Probability Rule:**
if $A \cup B = S$, where S is the sample space.
also, $A \cap B \neq \Phi$, then

$$P(A) = P(A \mid B) \times P(B) + P(A \mid \overline{B}) \times P(\overline{B}) \tag{6}$$

where, $P(\overline{B}) = 1 - P(B)$\{complementary probability\}

Bayesian rule helps to estimate the most probable underlying model for a random process, based on some observed data or evidence. If A represents a event and B represents the set of observed data values(evidence), then the terms in equation (5) are given the following terminology.

- P[A] = the prior probability of event A (in the absence of any evidence).
- P[B] = the posterior probability of B using the knowledge A.
- P[B|A] = the likelihood that the evidence B was produced, given that A was true.
- P [A|B] = the posterior probability of A being true, given that the evidence is B or B is observed.

From Bayes' Theorem we have,

if A and B are exhaustive events of sample space S, then :

$$\text{posterior} = \frac{\text{conditional likelyhood*prior}}{\text{likelyhood}}$$

where

conditional likelihood = P(A|B)

prior = P(B)

likelihood= P(A)

**Generalised Bayes Theorem:**

let $A_1, A_2, A_3 ........, A_{n-1}, A_n$ be mutually exclusive and exhaustive events. Then for any event B,

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)\, P(A_i) \tag{7}$$

\{total probability law\}

if $P(A_i) > 0$ for i=1,2....n , then for any event B where $P(B) > 0$ , we have the generalised Bayes theorem as:

$$P(A_k \mid B) = \frac{P(B \mid A_k)\, P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i) P(A_i)} \tag{8}$$

It is common to think of Bayes rule in terms of updating our belief about a hypothesis A in the light of new evidence B. Specifically, our posterior belief P(A|B) is calculated by multiplying our prior belief P(A) by the likelihood P(B|A) that B will occur if A is true.

The power of Bayes' rule is that in many situations where we want to compute P(A|B) it turns out that it is difficult to do so directly, yet we might have direct information about P(B|A). Bayes' rule enables us to compute P(A|B) in terms of P(B|A).

### 4.8.3 Application Variables

There are two basic amenities needed for any species to survive food and shelter. We take a simplistic model and only take food resources into consideration. In the following discussion the term 'region' refers to a sensor network patch as described in previous section.

Following is a list of mutually exclusive independent events will play an important role in deciding the densities of Cane toads within a region. The model starts with basic necessities, food and water. In future the model should add on more variables after constants of the simplified model are determined.

We base our model on three factors.

- Access to Water
- Food Resources
- Access to the region

A zone in the Kakadu National Park consists of points in a two dimensional space. For each point we define the Bernoulli Random variables $W_1$, $W_2$, $W_3$, $F_1$, $F_2$, $F_3$, $F_4$, $C_1$.

1. **Access to Water**. Measures effect of water resources.

   Water plays an important role in the toads life. The toad tries utmost to keep moist and preserve it's body water levels. The cane toads need water to lay their eggs. It has also been observed that the toads need to visit water once in three days. .

We therefore categorize the effects of water resources into three categories depending on how far water resources are: -

(a) **Immediate vicinity($\mathbf{w_1}$):** Refers to the water bodies which are with in a range of 1 km.

$$w_1 = \begin{cases} 0, & when\, no\, water\, resources\, in\, range\, of\, 0 \leq r \leq 1km \\ 1, & presence\, of\, water\, resource\, in\, range\, of\, 0 \leq r \leq 1km \end{cases}$$

where r is the radial distance of the water resource from the point.

(a) **Medium range**($\mathbf{w}_2$): Refers to the water bodies which are within a range of 1 km to 3 km but not less than 1 km.

$$w_2 = \begin{cases} 0, & when\, no\, water\, resources\, in\, range\, of\, 1 < r \leq 3km \\ 1, & presence\, of\, water\, resource\, in\, range\, of\, 1 < r \leq 3km \end{cases}$$

(a) **Far range($\mathbf{w_3}$):** Refers to the water bodies which are beyond 3 km and less than 10 km.

$$w_3 = \begin{cases} 0, & when\, no\, water\, resources\, in\, range\, of\, 3 < r \leq 10\, km \\ 1, & presence\, of\, water\, resource\, in\, range\, of\, 3 < r \leq 10\, km \end{cases}$$

There is no strong reasoning behind the figures of 1 km, 3 km and 10 km, considering the speed at which toad moves (27 km/year), we found it reasonable to assume that anything within 1 km is close and beyond 3 km is far.

2. **Food Resources**: Measures effect of food resources.

   Even though the cane toads attempts to eat anything smaller than itself, the main food for cane toads are the species below it in the food chain comprising of insects.

The main source of food for insects is vegetation, we have taken the liberty to assume that density of vegetation reflects the population of insects it can support. It may not be true in all cases but is true in most of the situations. Therefore, the denser the vegetation, the greater the insect population it can support and hence more food is available for the cane toads.
Similar to water we have categorised the food resources into four.

- $f_1$ - Dense Vegetation

$$f_1 = \begin{cases} 0, & when\, the\, vegetation\, is\, not\, dense \\ 1, & when\, the\, vegetation\, is\, dense \end{cases}$$

- $f_2$ - Mild Vegetation

$$f_2 = \begin{cases} 0, & when\, the\, vegetation\, is\, not\, mild \\ 1, & when\, the\, vegetation\, is\, mild \end{cases}$$

- $f_3$ - Other Vegetation(neither dense nor mild)

$$f_3 = \begin{cases} 0, & when\, other\, vegetation\, is\, not\, present \\ 1, & when\, other\, vegetation\, is\, present \end{cases}$$

- $f_4$ - Alternative Food Resources

$$f_4 = \begin{cases} 0, & when\, alternative\, food\, resources\, are\, not\, present \\ 1, & when\, alternative\, food\, resources\, are\, present \end{cases}$$

  Incorporates the effect of alternative food resources. Cane toads are highly competitive and adaptive species, they have been found to have stolen food from cat and dog food bowls. Alternative food resources also include species like ants and termites.

3. **Transport Corridors**

Dam et. al. [16] found that the spread of cane toads in the Kakadu National Park will depend on the transport corridors available to them. This includes pathways, roads, riverbeds and "hitch-hiking" in human vehicles. Observers on a field trip to study cane toads observed more toads along the roads than in the area of study.

Another factor that will affect the presence of cane toads in a region 'X' is the presence of cane toads in the adjacent regions. If the adjacent region has toads, given the transport it is highly probable that toads will occupy the region 'X', the opposite also holds true. Except the boundary regions, if adjacent regions do not have any toads it is unlikely that region 'X' will have toads.

$a_{i,j}$ - Measures effect of Access into zone 'i' from adjacent zone 'j'.

$$a_{i,j} = \begin{cases} 0, & when\, transport\, is\, not\, available\, from\, zone\, j\, to\, zone\, i \\ 1, & when\, transport\, is\, available\, from\, zone\, j\, to\, zone\, i \end{cases}$$

### 4.8.4 Intuitive Probabilities initialization

The inherent nature of the Bayesian approach is to provide rules that explain how one should change the existing assumptions or probabilities in light of new evidence. It allows to assign probabilities to known possible outcomes with unknown probabilities. These unknown probabilities change with observed outcomes.

An example is to imagine a new born kitten observe his first sunset and wonders weather sun would rise tomorrow or not. She assigns $\frac{1}{2}$ probability to sun rising the next day. The next day when the kitten observes sun rising, she updates the probability that sun will rise next day in future to $\frac{2}{3}$, thus increasing the belief that sun will rise the next day. The process is iterative and gradually the probability that sun will rise next day will creep closer to one, changing the belief with probability $\frac{1}{2}$ to a near certain event.

Bayesian Networks are also known as a Bayesian belief network. The higher the probability the more is the belief on a variable. For example, if $P(w_1)$ has the probability .8 and $P(w_2)$ has the probability .1 . This shows that there is higher probability of cane toad existing in region which is in immediate vicinity of a water body, and hence the higher belief on $w_1$.

Figure 17 shows the variables in our Bayes Net and figure 18 shows the Bayes Net that was built.

As discussed in the previous section, each location can be classified according to the food and water resources available around it. We define the following random variables for further simplification:-

W = presence of absence of water resources as an ordered triplet of $\{w_1, w_2, w_3\}$.

for example, $\begin{cases} 001 & for\, water\, resources\, from\, 3 < r \leq 10\, km \\ 000 & no\, water\, resources \end{cases}$

Note: At most one of the variables $w_1$,$w_2$,$w_3$can be initialised to one. This is due to the fact that if water resources are available in immediate vicinity then water resources in medium range will not affect the presence of cane toads in a location.

F = presence or absence of vegetation as an ordered triplet of $\{f_1, f_2, f_3\}$.

for example, $\begin{cases} 010 & for\, mild\, vegetation \\ 100 & for\, dense\, vegetation \end{cases}$

again, only one of the variables $f_1$, $f_2$, $f_3$can be initialized to one.

Using, W, F and $f_4$ we can generate all possible type of distribution of locations based on resources available. The distribution table below consists of the truth table listing of all the combinations of values of the system variables. The table can have 32 possible outcomes.

Each row in table 6 is a combination of values of W, F and $f_4$ and says how probable it is. The P(attraction) column shows the probability of finding cane toad at the particular location.

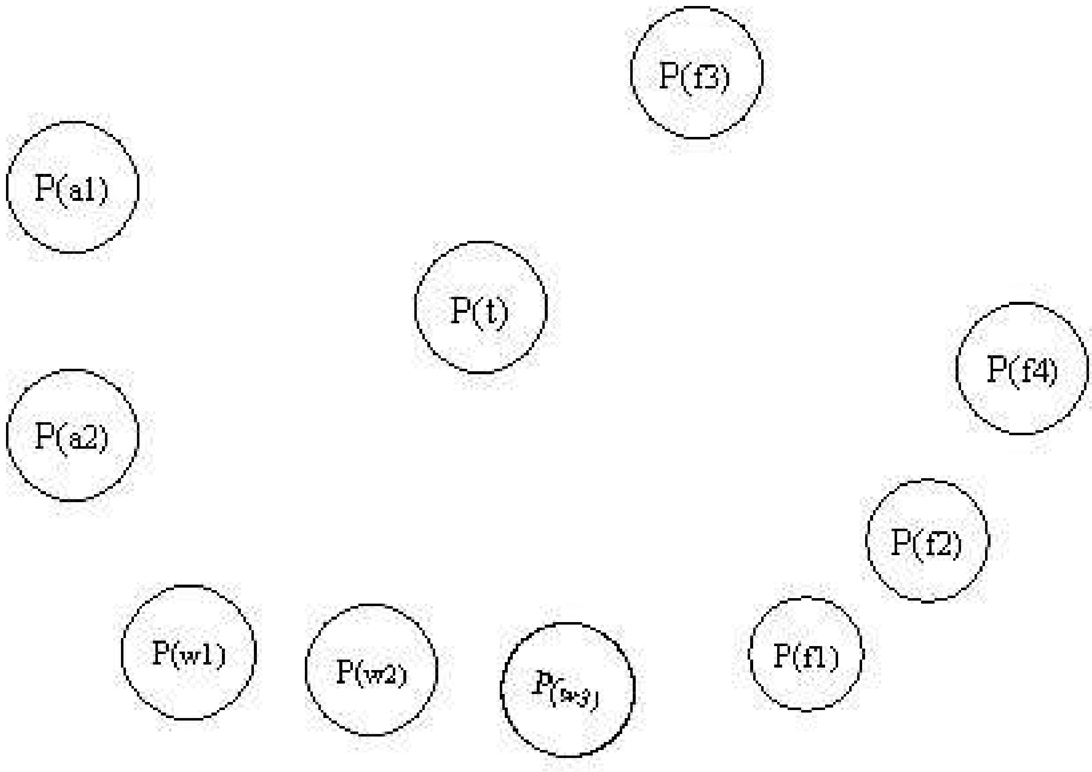


Figure 17: The Bayes Net variables

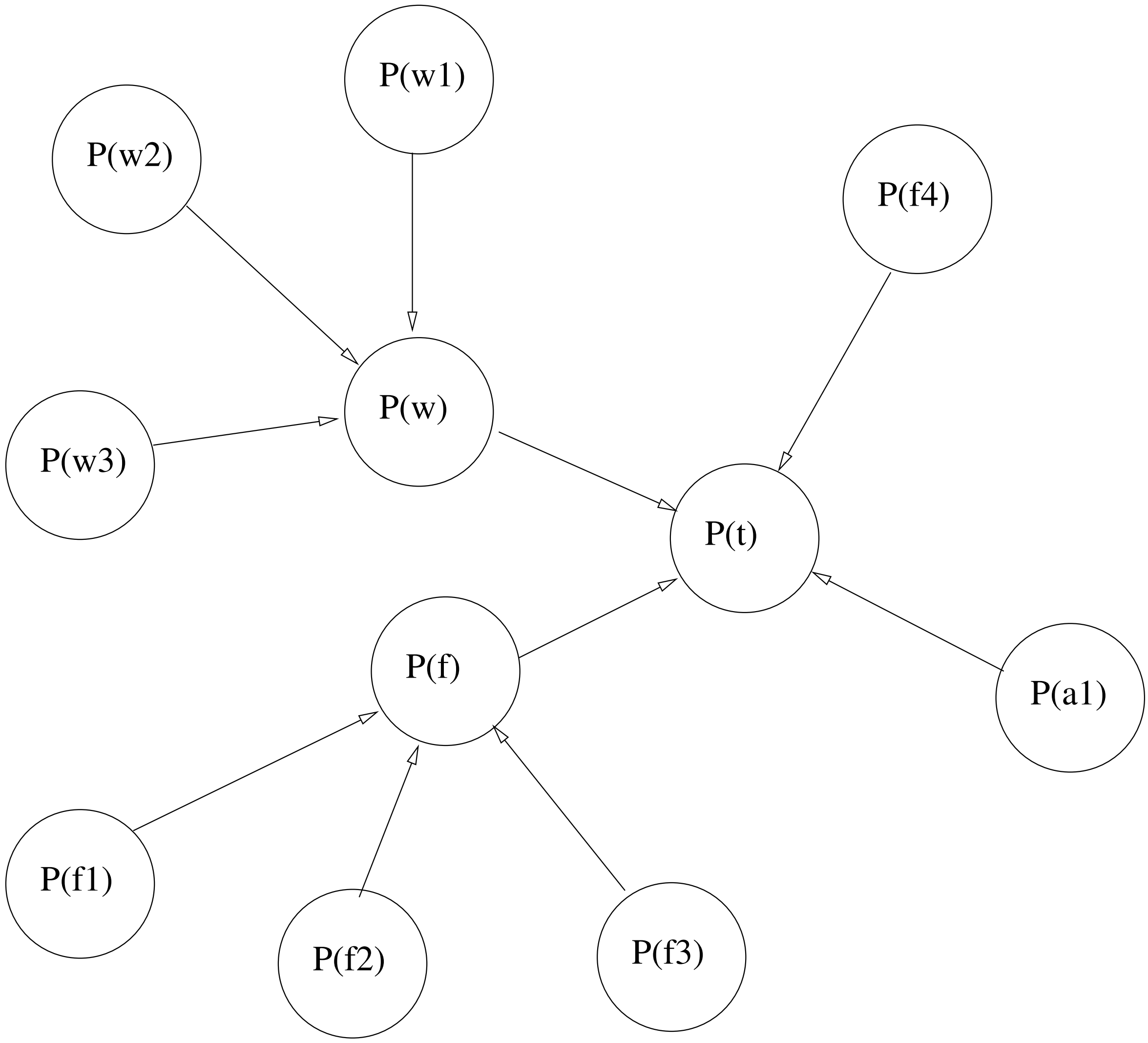


Figure 18: Bayes Network

| W | F | $f_4$ | P(attraction) |
|---|---|---|---|
| 000 | 000 | 0 | 0.01 |
| 001 | 000 | 0 | 0.11 |
| 010 | 000 | 0 | 0.05 |
| . | . | . | 0.15 |
| . | . | . | 0.21 |
| 100 | 010 | 1 | 0.14 |
| 100 | 100 | 1 | 0.31 |

Table 6: Distribution Table

We assume that the probability of finding a toad in any of the possible 32 outcomes (denoted by the rows in the distribution table) is the same. Also, the belief of each node (Figure 17) is assumed to be equally distributed. Each value of the variable has the same probability. We believe (for some time) that effect of water bodies and food resources is the same on the existence of cane toads.

$$X_i = Ordered\, triplet\, of\, \{W,\, F,\, f_4\} \qquad for\, i = 1\, to\, 32$$

(for all possible outcomes)

$P(X_0 = X_i) = \frac{1}{32} \qquad \forall_{i=1\, to\, 32}$

e.g. $P(X_0 = X_i = 000, 000, 0) = \frac{1}{32}$ where $(0 \leq P(X) \leq 1)$

The probabilities are updated when the sensor nodes detect the presence of cane toads at a particular location. As evidence is introduced to the network, the belief of the corresponding variable changes. This propagation effect is called belief update. The evidence is propagated through the whole network according to an algorithm that distinguishes inferior and superior nodes. The process is bottom up as the evidence is based on sighting of a toad.

If $n_t$ observations are are made till time 't' and $X=X_i$occurs $m_i$ number of times , {for i=1...32} then,

$$P(X_t = X_i) = \frac{1 + m_i}{32 + n}\, \forall_{i=1\, to\, 32}$$

also $\sum_{i=1}^{32} P(X_t = X_i) = 1$

The system therefore iteratively updates the probabilities on discovery of a evidence.

The above analysis was performed for a particular location in the park. We now extend it to predict the attraction level of a zone consists of finite number of such locations in the following section.

### 4.8.5 Prediction

We would like to predict the attraction level P(attraction) which denotes probability that a cane toad is found given the food and water resources.

Assume a zone 'j' denoted by $Z_j$, consisting of finite number of points, where $q_1$points are of type $X_1$, $q_2$points are of type $X_2$, ........, $q_i$points are of type $X_i$.

Let

T = Event that a cane toad is found.

P(t) is the probability that a cane toad is located in a region. Its measure of belief that cane toad will colonise the region given the resources.

Thus, the probability that a toad is found in a zone $Z_j$at a time 't' is

$$P(T \mid Z_j) = \frac{\sum_{i=1}^{32} q_i P(X_t = X_i)}{\sum_{i=1}^{32} q_i} \tag{9}$$

where the denominator($\sum_{i=1}^{32} q_i$) is the normalizing factor.

The above equation defines the the attraction level of a Zone 'j'. It can be observed that the attraction level is a function of P(t), P($w_1$), P($w_2$) , P($w_3$) , P($f_1$) , P($f_2$) , P($f_3$) and P($f_4$) and is reflected by the term $P(X_t = X_i)$.

Therefore the attraction level of a zone 'j' is

$$P_j(attraction) = P(T \mid Z_j) = \frac{\sum_{i=1}^{32} q_i P(X_t = X_i)}{\sum_{i=1}^{32} q_i}$$

We define a scope probability function which reflects the future scope of a region that is not yet habitated by the toads.

$P_j^{unoccupied}(attraction) = \sum_{i=1}^{i=all\ neighbours}(aP(attraction)_i + P(a_{i,j})) + bP(t \mid Z_j)$

where i refers to the adjacent node and a and b are weights. The first term takes the influence of neighboring regions into consideration. If the neighboring region is colonised by the cane toads the probability that this

region will be colonised is higher compared to the situation when none of the neighboring nodes have any cane toads.
$P(a_{i,j})$ refers to the access routes available from zone 'i' to zone 'j'.

Therefore

$$P_j^{occupied}(attraction) = \frac{\sum_{i=1}^{32} q_i P(X_t = X_i)}{\sum_{i=1}^{32} q_i} \tag{10}$$

$$P_j^{unoccupied}(attraction) = \sum_{i=1}^{i=all\,neighbours} (aP(attraction)_i + P(a_i)) + bP(t \mid Z_j) \tag{11}$$

This future scope prediction holds true only for regions which do not have toads. Once the toads occupy a region, to a large extent the dynamics are determined by the local factors instead of the ergonomics of the adjacent regions.

Scope function assists in predicting the direction of growth and movement of cane toads. It will also help selectively deploying sensor networks in regions with high scope and ignoring regions with less scope.

### 4.8.6 Seasonal Influence

So far we have ignored the influence of the seasons. Seasons play a big role in cane toad's activity. Cane toads are more active in the wet seasons of the year and relatively dormant in the dry seasons when they are trying to conserve their water resources. A lot of temporary water bodies like mud-holes are created in wet seasons which need to be taken into account. Alternatively dry seasons might make a region less habitable due to lack of water bodies.

To account for the effect of seasons we multiply the attraction level P(attraction) with a seasonal effect multiplying factor.

Let $s_{Z\ S}$ be the multiplying factor for region 'z' in season 's'

$s_{z\,dry} < s_{z\,wet}$ as wet seasons enhance the attraction level of a zone.

Kakadu has seasons of varied extremes. The park's aboriginal inhabitants have divided the year into six distinct seasons. [44]

- Gudjewg (January, February):Violent thunderstorms, heavy rain and flooding.
- Banggereng (March): Expanses of water recede and streams run clear.
- Yegge (April, May): Drying winds, bush fires.
- Wurrgeng (June, July): Cold weather and low humidity.
- Gurrung (August, September):Gurrung is windless and hot.
- Gunumeleng (October, November, December): Pre-monsoon season of hot weather, which becomes increasingly humid.

The value of $s_{Z\ S}$ can be estimated from the Bayes Net. It can be measured from the number of toads observed in different seasons at the same location relative to a season which is taken as a base metric.

The attraction level of a zone measured in previous section is adjusted by a factor $s_{Z\ S}$. We define a 2 dimensional matrix. The matrix has season as rows, and $s_{z\,s}\ P(attraction)$ as the columns where P(attraction) is derived from equation (10) and (11).

| *gudjewg* | 1 | 2 | 3 | $z_i$ | ... | $z_{n-1}$ | $z_n$ |
|---|---|---|---|---|---|---|---|
| *banggereng* | .23 | .01 | .95 | .001 | .23 | .12 | .54 |
| *yegge* | .28 | .54 | .256 | .012 | .28 | .95 | .177 |
| *wurreng* | .29 | .177 | .644 | .112 | .29 | .256 | .687 |
| *gurrung* | .37 | .687 | .121 | .45 | .37 | .644 | .001 |
| *gunumeleng* | .34 | .12 | .101 | .75 | .34 | .121 | .101 |

n = total number of nodes.

$\bigtriangledown i\ \{i\,|z_i\,\varepsilon ALL\,ZONE\ \wedge 0 < i \leq n\}$

In the above matrix, 1, 2, 3, 4 .. $z_n$ are are the zones.

The matrix can be used to find the maximum P(attraction) for a zone and then sensor nodes can be deployed in that zone. For best performance, the deployment node should be based on the maximum value of P(attraction) . We can also use to predict the requirements in a given season for a zone, or what zone will be the hotspot in coming season . It can also be used to predict the next zone that is more likely to be colonized.

## 4.9 Summary

The geographical region of Kakadu National Park is completely explored (figure 19), and statistical data like water bodies, type of vegetation at a particular location and roads is well documented. This information will be input to the system. A sensor node will use the location information to find the environment information (water and vegetation).

We assume with $P(w_1)$, $P(w_2)$, $P(w_3)$ are exclusive events to a large extent. There might be a possibility that presence of a water body in medium range might enhance the probability in immediate vicinity, we are ignoring this effect. Likewise, influence of water bodies on the vegetation in a geographical location have been ignored.

A possible error which can be induced due to the linear model of perceptron learning is that we have assumed linearity in the equation which might not be close to the truth. It is a prediction based model and more work should be done to improve it. It does not take into effect the time cane toads have colonised a region, As toads increase in a region the resources available decrease and therefore the region's attraction level goes down. Localization techniques used would help predict the direction of movement and update the variables in the adaptive learning algorithm.

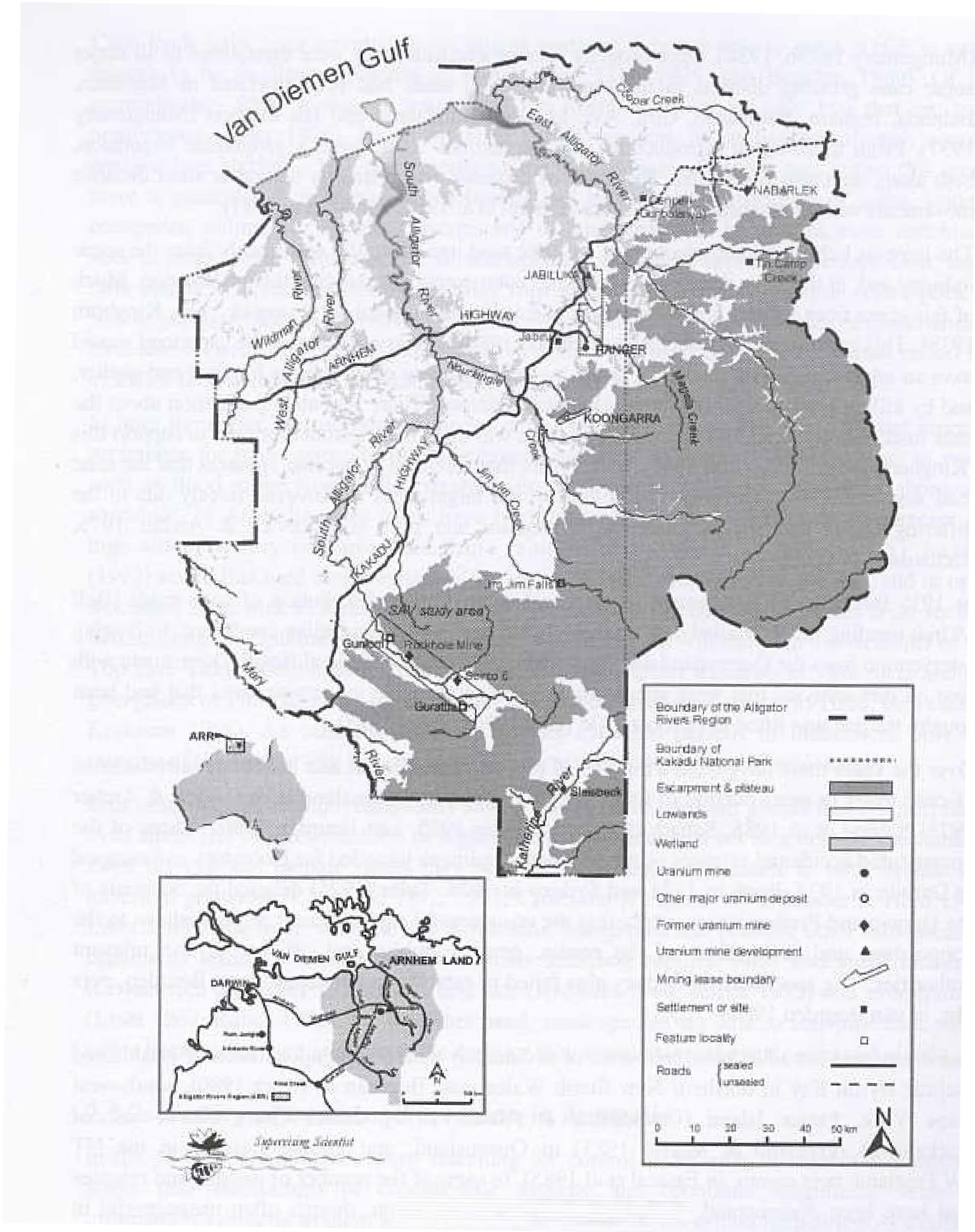


Figure 19: Map of Kakadu National Park

# 5 Evaluation Strategies

It is hard or often unrealistic to judge how well a deployment scheme will work without deploying it first. In this chapter we discuss the potential deployment metrics to measure the success of the deployment scheme in the zones. The deployment should be able to satisfy the functional requirement of the system. The potential metrics should be able to measure how well the functional requirements are satisfied.

The two fundamental requirements of the habitat monitoring system are target detection and target localization. The first is to determine whether observed animal calls belong to cane toads using their spectrograms. The second is to locate the position of cane toad when its vocalization is recognized.

Furthermore we discusses the various metrics that would help determine the performance and success of the deployed sensor network.

## 5.1 Potential Metrics

In this section we discuss the various metrics that define the Quality of Service(QoS) for our sensor networks application. In sensor networks the quality of service is determined by the system requirements. For example using sensor networks for forest fire detection one may ask how well the network can observe a given area and what the chances are that a fire starting in a specific location will be detected in a given time frame.

### 5.1.1 Ease of deployment

The network will contain hundreds of thousands of nodes, the deployment should not require any individual setup or placement. The network is expected to function unattended and untethered. The sensor nodes are not expected to run forever, the nodes once deployed might die due to battery drainage or other natural causes. In such cases additional sensor nodes will have to be deployed to maintain the threshold exposure and detection accuracy. The number of times node replenishment is required and the mode of deployment and redeployment in the sensor patches will increase the deployment cost and the maintenance cost of the system. These costs are not governed by the number of sensor nodes required but by the human resources required for the effort. The more human intervention is required the higher will be the cost.

### 5.1.2 System Lifetime

The system lifetime is the time the system once deployed functions unattended, smoothly and continuously while satisfying all the functional requirements. The sensor nodes are deployed in high densities and therefore if a few nodes die, it does not affect the system dynamics drastically. In a few cases it many not have any affect at all. In our case the system lifetime is the time until the probability of missed detection or false detection exceeds a threshold. An interesting thing to study would be how much can we afford to reduce quality of service of sensor network to increase the system lifetime.

### 5.1.3 Coverage

Coverage in general is the quality of service that can be delivered by a sensor network deployment. The coverage required can be measured as a best case coverage or a worst case coverage. For example we can have worst case coverage as a metric for one system that does not require high detection range and best case coverage as a metric for a different system which has high performance requirements. Worst case coverage, measures the quality of service by finding areas of lower observability from sensor nodes and detecting weak regions. Best case coverage requires finding areas of high observability.

The geometry of Kakadu (Figure 19) is known in advance and is input to the system for the adaptive Bayesian Net algorithm. By combining computational geometry and graph theoretic techniques like graph search algorithms, optimal worst, best and average case for coverage can be calculated. Meguerdichian et. al. [45] gave polynomial time centralized algorithms to solve the questions optimally. However, their algorithms rely heavily on some geometrical structures such as the Delaunay triangulation and the Voronoi diagram (Figure 20) which cannot be constructed locally or even efficiently in a distributed manner.

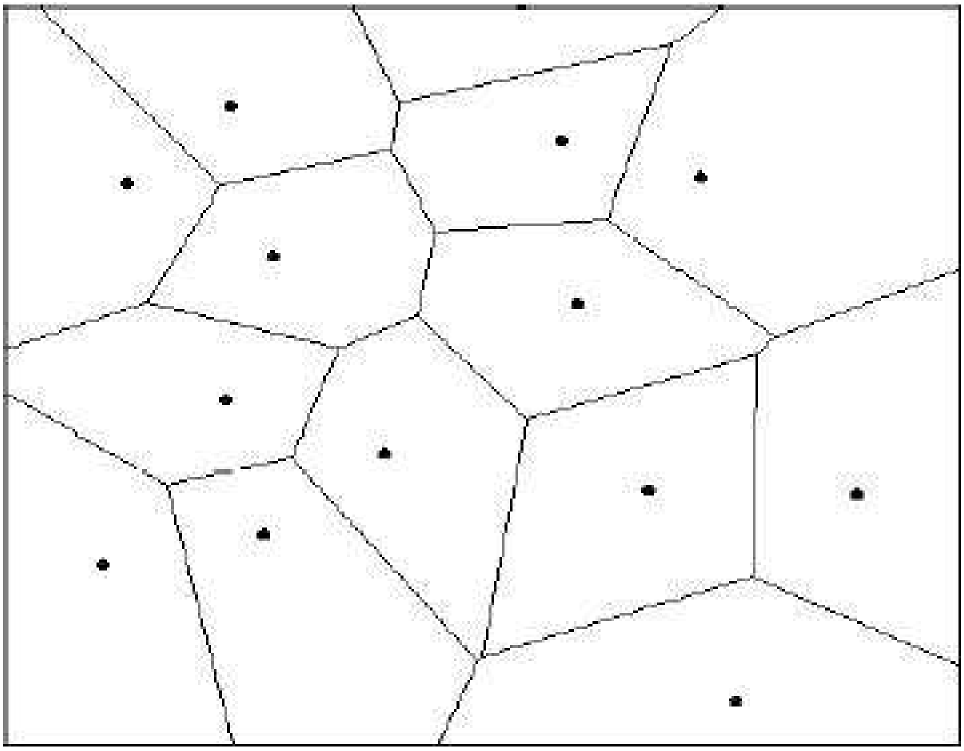

Figure 20: Voronoi Diagram

The optimum coverage strategy should be the one where the path between the cane toad and its closest distance to any of the sensors is as large as possible without compromising the detection accuracy. This would reduce the bandwidth requirement, numbers and density of sensor nodes deployed and the cost. The equation to calculate the coverage is given in section 4.7.7

### 5.1.4 Exposure

Exposure is a measure of how well an object, moving on an arbitrary path, can be observed by the sensor network over a period of time. Figure 21 shows an example of exposure, the arc in the figure is the path the object traverses. Exposure is directly related to coverage, the more the coverage the better the exposure could be.
It is hard and complex for the system to track each and every toad as vocalization techniques are species centric and not individual centric. Individual centric vocalization would require complex speech recognition software which

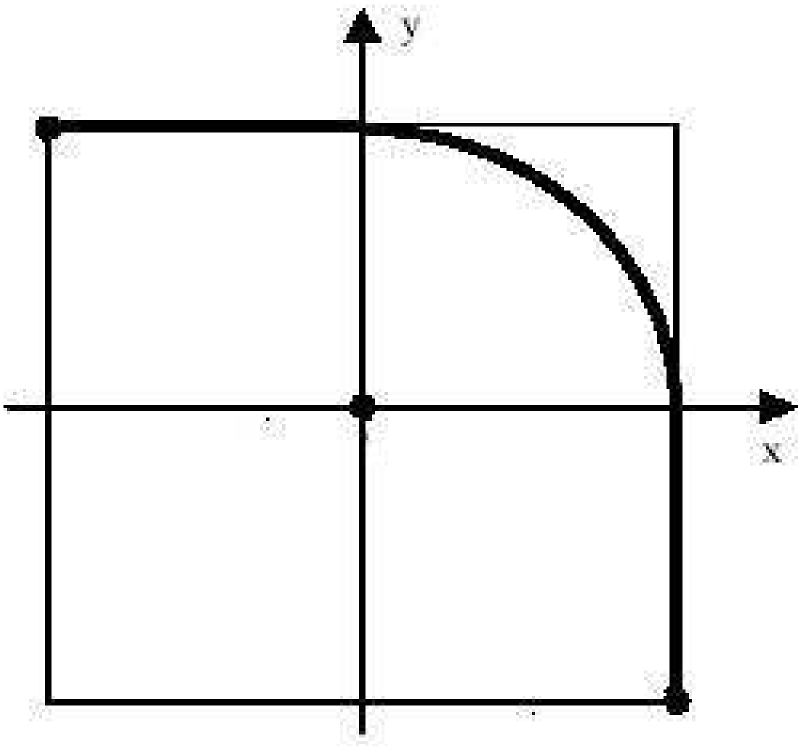


Figure 21: Example of Exposure

is hard to run in a sensor network environment considering the energy and processing capabilities that will be required.

Exposure metrics should not be used to track individual toads instead they should try to determine the directional movement and growth of cane toads from one region to another. This information should be used to analyze how well the Adaptive learning algorithm performs. The adaptive learning algorithm attempts to predict the regions "hotspots" where the probability of cane toad colonizing a zone is high.

#### 5.1.5 Detection Accuracy

Detection accuracy means given an animal call how well the system is able to accept or reject it as a cane toad's vocalization. In the ideal case, we would like to have 100% detection accuracy. This is however not possible as significant amounts of noise are present in some regions from insects, rain, birds, human activity and other animal calls.

In general the detection accuracy depends on SNR (Sound to Noise Ratio). Noise from wind, woods, brushes usually has a different frequency from that of animal calls because the evolution favors species that can make themselves heard clearly.

The voice recognition system will filter out the noise from winds, rains, bushes. The noise that matters is that in the same bandwidth as the animal calls. The system might fail due to the presence of a large different species of frogs. Frogs and toads have evolved from a common species, and therefore they have similar vocalizations, their active times are similar to the cane toads. A large amount of noise is expected in the form of frog vocalizations.

It is preferable for the system to fail to recognize a vocalization (a false negative) than to incorrectly indicate the presence of cane toads (a false positive). It is critical to choose thresholds such that false positives are unlikely. It is easy to provide large amounts of training data which only contains the vocalizations of a cane toads however often it is hard and time consuming to manually extract the vocalizations that positive training data requires [12].

### 5.1.6 Cost

The final metric that determines how well an application has performed is the cost incurred to get the desired gains. The cost incurred can be direct and indirect. Direct cost refers to the expenses pertaining to sensor nodes, base stations, macronodes(Plebs) [46], batteries etc. Indirect costs would refer to some thing like the number of man hours required to configure, monitor, deploy, redeploy the sensor network.

For sensor platforms cost and network lifetime are directly proportional to the energy. Communication is usually the main source of energy dissipation in sensors, which greatly depends on the distance between the source and destination of a communication link. Clustering, specifically in sensor networks, could be used to solve a variety of problems. Clustering can be used for intra-cluster aggregation to reduce the amount of communication between the sensor nodes. A clustering approach similiar to Heinzelman et. al. [14] could be used to transmit processed data to base stations, hence minimizing the number of nodes that take part in long distance communication.

The driving force behind the deployment cost will be the macronodes(Plebs). The micronodes (sensor nodes) are expected to cost a few cents and the macronodes (plebs) will cost a few hundred dollars.

Due to the channel capacity, there is a maximum number of sensors that each macronode can handle. This does not allow each sensor to communicate to its closest macronode, as that macronode might have already reached its service capacity. This however is not the deciding factor in determining the number of plebs required.

The number of plebs required will be determined by the bandwidth required to transmit the acoustic data over the sensor nodes. The bandwidth required is constrained by the number of sensors nodes. The sensor nodes act as routers and there is an upper limit to the data rate they can support. Sensors send data to the macronodes over a multi hop wireless network arranged in a hierarchical structure. The more the hops the higher will be the data rate required. For a hexagonal cell structure (figure 22), it was calculated that the data rate required increases exponentially with every hop towards the macronode. The data rate is propotional to $2^m$ where m is maximum number of hops to the pleb .

The number of hops 'm' is directly proportional to $\frac{Number\, of\, Nodes}{Number\, of\, Plebs}$
let m = number of maximum hops the network can physically support.

to the pleb is given by:

$$Total\, Nodes\, per\, pleb = 6(2^{m-1} - 1)$$

Let

n = number of sensor nodes

k = number of plebs

$m$ =maximum number of hops to a pleb from a sensor node

**MicroNodes Requirement Calculation:-**

Assuming a hexagonal cell structure:

d - diameter of the hexagonal cell = 20 m .

Area covered by one micronode = $\frac{3\sqrt{3}d^2}{2}$ = 1040 $m^2(approx)$

Area of a Zone = $10^7 m^2$

Number of sensor nodes required = $\frac{Area\, of\, zone}{Area\, covered\, by\, one\, micronode}$ = 9600 nodes

Node density = 9.6 x $10^{-4}$ $\frac{nodes}{m^2}$

**Bandwidth Requirement:-**

The highest vocalizations are of frequency 8 KHz.

with 8 bit sampling, the data rate = 128 kbps

using compressing algorithm with S-encoding compression of less than 1% can be acheived [32].

therefore, data rate required at sensor node for one vocalization = 1.28 kbps

assuming, maximum data rate a sensor node can handle $\simeq$128 kbps.

In a worst case scenario, the maximum simultaneous vocalizations though distinct micro sensor nodes that can be supported $\simeq$ 100

**Pleb Requirement:-**

The maximum expected activity of the toads is expected to be in the night when they cluster together and emit
their mating calls. This would mean a set of micro sensor nodes in few locations that are near to a water body will receive incessant vocalizations while others will be idle. Assuming 5% of the micro nodes are in this high activity region.

The number of sensor nodes under a pleb = $\frac{Worst\, Case}{percent\, of\, active\, nodes}$ = 2000

The number of plebs required for a zone = $\frac{Total\, Number\, of\, Nodes}{Number\, of\, Nodes\, per\, pleb}$ =$\frac{9600}{2000}$ $\simeq$ 5

**Cost Estimates :-**

k = number of plebs required

| | Cost of Sensor Nodes | Cost of Plebs | Total Cost |
|---|---|---|---|
| **Using Only Plebs** | 0 | $4.8 \times 10^6$ | $4.8 \times 10^6$ |
| **Using Only Senor Nodes** | 2400 | 0 | 2400 |
| **Using our proposed approach** | 2400 | 2500 | 4900 |

Table 7: Cost estimates for a zone (in US Dollars)

| | Cost of Sensor Nodes | Cost of Plebs | Total Cost |
|---|---|---|---|
| **Using Only Plebs** | 0 | $8680 \times 10^6$ | $8680 \times 10^6$ |
| **Using Only Senor Nodes** | $4.32 \times 10^6$ | 0 | $4.32 \times 10^6$ |
| **Using our proposed approach** | $4.32 \times 10^6$ | $5 \times 10^6$ | $9.32 \times 10^6$ |

Table 8: Cost estimates for the Kakadu National Park (in US Dollars)

Assuming a sensor node costs $c_1$ and pleb cost $c_2$

Direct Cost $= n \times c_1 + k \times c_2$[1]

90% of Kakadu can be covered by sensor nodes, factoring the 90% we estimate the cost of instrumenting the entire Kakadu National Park.

Tables 7 and 8 show the relative cost estimates for instrumenting one zone and the whole park with only plebs, only sensor nodes and combination of plebs and sensor nodes in a clustered environment respectively. The final cost will much less due the hotspot identification by the adaptive algorithm proposed.

## 5.2 Potential Methodology

It is impossible to measure the success of deployment from the adaptive learning algorithm, because one will never exactly know where all the cane toads are. The system is constrained by localization accuracy and individual toads can not be tracked. The adaptive learning algorithm is dependent on the number of toads that are detected. We cannot detect any more toads than the number of toads the system can detect i.e. One of the constrains of the system is the system itself.

The location tracking and detection accuracy should be simulated and tested under laboratory conditions. The errors in localization and the detection failure rates should be factored before making any predictions from

[1]Cost of a pleb and a sensor node are estimated to be 500$ and 25 cents respectively.

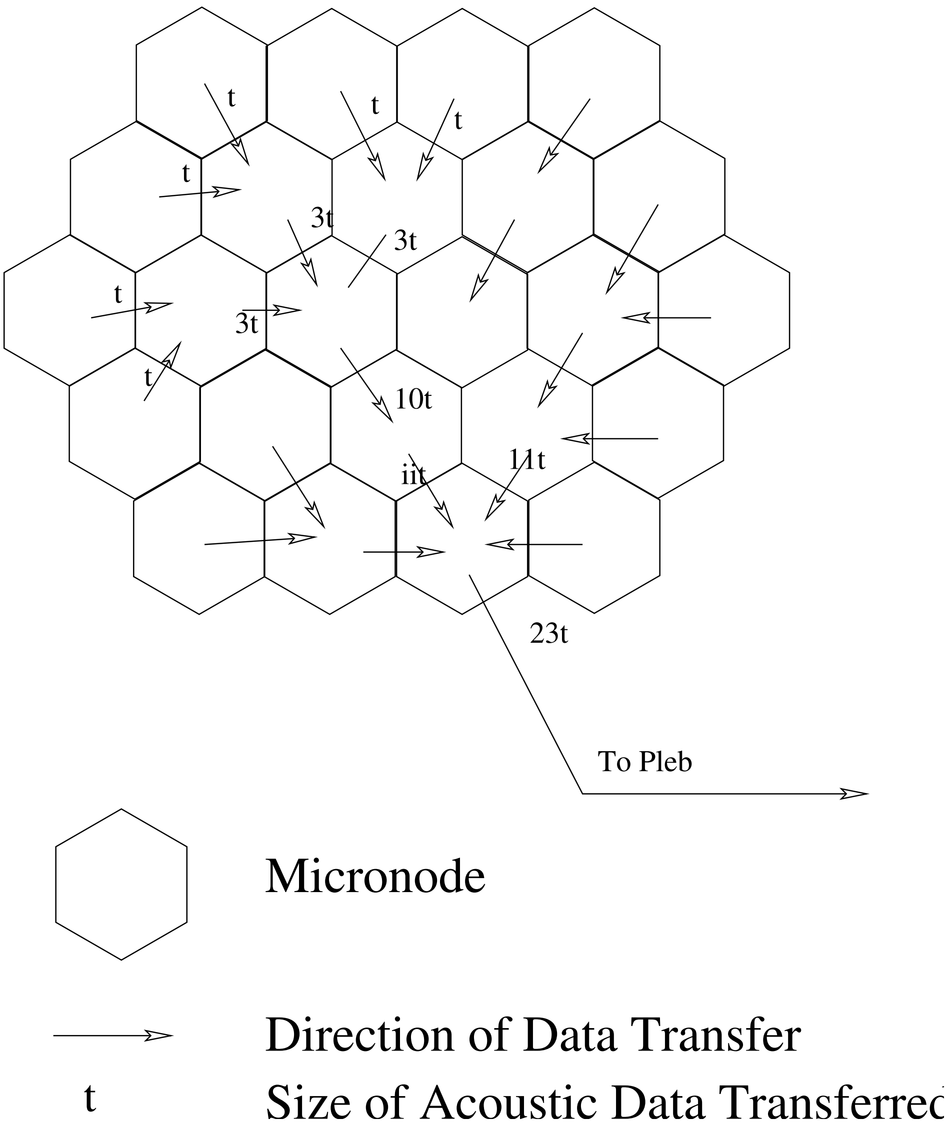


Figure 22: Data Transfer along the micronodes

the adaptive learning algorithm. The prediction of the adaptive learning algorithm should be compared with the results obtained from exposure and detection analysis to measure how well the algorithm predicted the hotspots.

For frogs, the highest frequency with non-trivial energy seldomly goes beyond 8 KHz, and the frequency usually spreads in a width of a few KHz. This would require a data rate of 128 kbps with 8 bit samples. Even though the frequency is for frogs, cane toads will have similar vocalization. The data rate of 128 kbps is high for a sensor node to handle. Grigg et. al. [17] monitored frogs using the plebs, their experiments show that 15 seconds of sound takes approximately 1 minute to process on a 25 MHz Intel 486 CPU.

The system therefore needs to be optimized for processing and communication overheads. Experiments should be performed to measure the bandwidth needed for acoustic data and possible data reduction and compression algorithms at the micronodes should be used. Wang et. al. [32] have utilized data reduction and compression techniques like S-coding that compactly encodes reduced acoustic signals. Their results show that 256,000 bits of waveform can be compressed to 3,064 bits by compressing the S-coded wave forms.

## 5.3 Summary

In this section we discussed the various metrics that would help determine how successful the deployment strategy has been. The metrics discussed were coverage, exposure, detection accuracy and cost. Coverage and exposure metrics will not only help in determining how well the sensor network satisfied the functional requirements but also how well the adaptive learning algorithm predicted the hotspots . We gave an analysis of the cost metric for cane toad monitoring application and discussed the potential methodology to highlight the simulation and experiments that need to done.

# 6 Future Directions and Conclusion

## 6.1 Future Directions and Challenges

A big challenge is to adapt the application to abrupt or slowly changing environmental characteristic with time, for example seasonal influences like wet seasons when toads are highly active and breed compared with dry seasons when they go into hiding to prevent the loss of water from the body.

The coverage formulations should try to find weak points in a sensor field and suggest future deployment or reconfiguration schemes for improving the overall quality of service. Coverage should sufficient enough to be able to track presence of all cane toad populations correctly.

For target detection and localization obstructions to the line of sight is a critical problem. Acoustic ranging performance suffers when the line of sight path is obstructed. Acoustic range measurements in obstructed conditions often consistently detect longer reflected paths leading to error.

The overhead for cane toad tracking and detection should be minimized to boost system life time. If we examine the dynamics, as working nodes die and sleeping nodes start working, the working node topology changes will consequently alter the coverage, exposure and detection.

Coverage should be simulated under laboratory condition. It is is easier simulate coverage compared to detection accuracy which will largely depend of the dynamics of Kakadu, i.e. rain, sound from wind, bushes, frogs and other animal calls.

The adaptive learning algorithm is based on primitive requirements like food and water. It can be improved by considering complex ecosystems within Kakadu. The overall environment of an ecosystem might be a better measure than primitive requirements. The ecosystems that can be considered include coastal plains, flood plains, lowland plains. The model should also consider the time cane toads have colonized a zone, after some time the resources of a zone will reduce and the zone will be lesser hotspot than it was earlier. Effects of landforms like escarpment, plateau should also be factored.

Several researchers have tried to map the distribution of the cane toads with the climatic conditions available, across all the set of regions where cane toads have been found. These models have successfully predicted the presence of cane toads on a macro level. Few of the input parameters of the model like the soil moisture retention capability, temperature and stress conditions can be easily incorporated into the adaptive learning algorithm.

## 6.2 Conclusions

In this case study, we present a multi variate problem of monitoring the cane toads using a sensor network and propose a novel framework for sensor network deployment by integrating application, economic, and networking/technology objectives for habitat monitoring.

Deployment has been analyzed in terms of a higher-level application perspective when many objectives have to be satisfied simultaneously. Deployment problem was reduced into zone classification, in-zone deployment. Since it is hard to get initial deployment right due to uncertainties in the problem domain, an adaptive learning method using Bayesian Net was proposed and using it to identify the zones that might be "hotspots" for the cane toads and drive the deployment plan.

The adaptive learning algorithm, handled the uncertainties by making some initial probability assumptions and updating probabilities upon discovery of some evidences. The model considered the effects of water bodies, food resources, neighbouring zones, transport routes and seasonal influences to measure the feasibility of that zone being colonized by cane toads.

For the in zone deployment, we choose the hierarchical organization of sensor networks in order to reduce wireless communication and thus energy consumption by distributing signal processing to local micronodes and clusters. The in zone-deployment is a hierarchical hybrid network of possibly many mutually disconnected clusters. The clusters consists of micronodes and macronodes. The micronodes are sensor nodes and the macronodes (plebs) act as cluster heads as they have more energy, memory and computational capabilities. Clustering minimizes the number of sensor nodes that take part in long distance communication.

We also proposed metrics and deployment heuristics to measure the performance of learning algorithm and how well the functional specifications were serviced by the deployment.

Even though 90% of Kakadu is habitable by the Cane toads, they are expected to concentrate in a few regions which offer the better food and ecosystem for their survival. The adaptive learning algorithm will help in selectively identifying these regions which are probable hotspots and bringing the cost down.

**References**

[1] T. Camp, J. Boleng, and V. Davies, *A Survey of Mobility Models for Ad Hoc Network Research*, Wireless Communication & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications, vol. 2, no. 5, pp. 483-502, 2002

[2] C. Bettstetter. *Mobility Modeling in Wireless Networks: Categorization, Smooth Movement, and Border Effects*, ACM Mobile Computing and Communications Review, Vol. 5, No. 3, July 2001.

[3] Simon, H. A. *Why should machines learn* ?

[4] Leake, David B. *Artificial Intelligence*, [Online] Available: http://www.cs.indiana.edu/%7Eleake/papers/p-01-07/p-01-07.html, [October 2003]

[5] Luger, G. 2001, *Artificial Intelligence* , Pearson Education

[6] Tian, J. and J. Pearl (2000). *Probabilities of Causation: Bounds and Identification.* Annals of Mathematics and Artificial Intelligence 28, 287-313.

[7] Qi H., Kuruganti P. T.and Xu Y., *The Development of Localized Algorithms in Wireless Sensor Networks* [Online] Available: http://www.mdpi.net/sensors/papers/s20700286.pdf [October-November 2003]

[8] Hof Robert D., *The quest for next big thing*, [Online] Available: http://www.businessweek.com/magazine/content/03_34/b3846618.htm [October 2003]

[9] K. Sohrabi, W. Merrill, J. Elson, L. Girod, F. Newberg, and W. Kaiser, *Scalable self-assembly for ad hoc wireless sensor networks*, in Proc. IEEE CAS Workshop on Wireless Communication and Networking, Pasadena, CA, USA, September 2002.

[10] L. Girod, V. Bychkovskiy, J. Elson, and D. Estrin, *Locating tiny sensors in time and space: A case study*, in Proc. ICCD02, Freiburg, Germany, September 2002.

[11] Wang H.,Elson J., Girod L., Estrin D., Yao K., *Target Classification and Localization in Habitat Monitoring.* [Online] Available: http://lecs.cs.ucla.edu/Publications/papers/Wong03Target.pdf [September 2003]

[12] Grigg G., McCallum H., Taylor A., Watson G. *Monitoring Frog Communities: An Application of Machine Learning* [Online] Available: http://www.cse.unsw.edu.au/~andrewt/papers/frog_iaai96/frog_iaai96.html[August 2003]

[13] K. Yao, R.E. Hudson, C.W. Reed, D. Chen, and F. Lorenzelli. *Blind beamforming on a randomly distributed sensor array system*, [Online] Available: http://www.janet.ucla.edu/awairs/download_pubs/ kyao.JSACoct98.pdf [August 2003]

[14] Heinzelman, W. R.; Chandrakasan, A.; Balakrishnan, H. Energy efficient *Communication Protocols for Wireless Microsensor Networks*, Proc. Hawaaian International Conference. on Systems Science, 2000.

[15] *Cane toads getting the jump on native frogs* [Online] Available: http://www.smh.com.au/articles/2003/06/06/1054700384033.html [September 2003]

[16] RA Van Dom, DJ Walden and GW Begg 2002. *A preliminary risk assesment of cane toads in Kakadu National Park.* Scientist Report 164, Supervising Scientist, Darwin NT.

[17] Grigg G., McCallum H., Taylor A., Watson G. Monitoring Frog Communities: An Application of Machine Learning [Online] Available: http://www.cse.unsw.edu.au/~andrewt/papers/frog_iaai96/frog_iaai96.html[August 2003]

[18] *Exploratory Research*, [Online] Available:

http://www.intel.com/research/exploratory/heterogeneous.htm [October 2003]

[19]*μAmps Project* [Online] Available http://www-mtl.mit.edu/research/icsystems/uamps/uAMPS-2/index.html, [October 2003]

[20] *Millennial Net*, [Online] Available http://www.millennial.net/content.cfm?section=1 [September 2003]

[21] *Texas Instrument*, [Online] Available http://www.futurlec.com/News/TI/NewMicro.html, [September 2003]

[22] Joseph Robert Polastre, *Design and Implementation of Wireless Sensor Networks for Habitat Monitoring Masters Thesis.* [Online] Available:[October 2003]

[23] Vlado H., Andreas K., Holger K., Adam W. *A common wireless sensor network architecture?* [Online] Available http://www.tkn.tu-berlin.de/publications/papers/sensornetz_architecture.pdf . [September 2003]

[24] Howard A., Mataric M. J., Sukhatme G. S., *An Incremental Self-Deployment Algorithm for Mobile Sensor Networks*, [Online] Available: http://www-robotics.usc.edu/~ahoward/pubs/howard_ar02a.pdf [November 2003]

[25] Estrin, D., Govindan, R., Heidemann, J., Kumar, *S. Next century challenges: scalable coordination in sensor networks.* Intl. Conf. on Mobile Computing and Networking (MobiCom) 1999, 263-270.

[26] Intanagonwiwat, C., Govindan, R., Estrin, *D. Directed diffusion: a scalable and robust communication paradigm for sensor networks.* MobiCom 2000.

[27] Silva, F.; Heidemann, J., Govindan, R. *Network Routing Application Programmer's Interface (API) and Walk Through 9.0. [*Online] Available

http://www.isi.edu/~johnh/PAPERS/Coffin01a.html [November 2003]

[28] Kulik, J.; Heinzelman, R. B.; Balakrishnan, H. *Negotiation-based protocols for disseminating information in wireless sensor networks.* Submitted to ACM Wireless Networks 2000.

[29] W. Heinzelman, J. Kulik, and H. Balakrishnan, *Adaptive Protocols for Information Dissemination in Wireless Sensor Networks*

[30] Hollick M., Krop T., Schmitt J., Huth H., Ralf S., *Modeling mobility and workload for Wireless Metropolitan Area Network* [Emailed by the authors]

[31] Murphy K., *A Brief Introduction to Graphical Model and Bayesian Networks* , [Online] Available http://www.ai.mit.edu/~murphyk/Bayes/bayesrule.html [October 2003]

[32] Wang H., Estrin D., Girod L., *Preprocessing in a Tiered Sensor Network for Habitat Monitoring* [Online] Available http://lecs.cs.ucla.edu/Publications/papers/tier.pdf[September 2003]

[33] N.Bulusu,D.Estrin,L.Girod,and J.Heidemann,*Scalable coordination for wireless sensor networks:self-configuring localization systems*,in Proc.6th International Symposium on Communication Theory and Applications (ISCTA 01).

[34] L.Girod and D.Estrin,*Robust range estimation using acoustic and multimodal sensing*,in Proc.IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2001),Maui,Hawaii,USA,October 2001.

[35] J.Elson and D.Estrin,*Time synchronization for wireless sensor networks*,in Proc.2001 International Parallel and Distributed Processing Symposium (IPDPS),Workshop on Parallel and Distributed Computing Issues in Wireless and Mobile Computing ,p.186,San Francisco,Calif,USA,April 2001.

[36] A.Mainwaring,J.Polastre,R.Szewczyk,D.Culler,and J.Anderson,*Wireless sensor networks for habitat monitoring*,in 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA 2002),Atlanta,Ga,USA,September 2002.

[37] *Untited*, [Online] Available http://www.pc104.org/consortium/ [November 2003]

[38] K.Sohrabi,W.Merrill,J.Elson,L.Girod,F.Newberg,and W.Kaiser, *Scaleable selfassembly for ad hoc wireless sensor networks*,in Proc.IEEE CAS Workshop on Wireless Communications and Networking ,Pasadena,Calif,USA,September 2002.

[39] Allen G., *What Is Beamforming* [Online] Available http://www.ece.utexas.edu/~allen/Beamforming/ [November 2003]

[40] Min R., Bhardwaj M., Cho S., Shih E., Sinha A., Wang A., Chandrakasan A., *Low Power Wireless Sensor Low Power Wireless Sensor Networks*, [Online] Available http://www.mit.edu/~rmin/research/min-vlsi01.pdf [November 2003]

[41 ]K. Chakrabarty, S. S. Iyengar, H. Qi and E. Cho, *Grid coverage for surveillance and target location in distributed sensor networks*, [Online] Available http://aicip.ece.utk.edu/publication/02tc.ps [November 2003]

[42] Thomas Clouqueur, Veradej Phipatanasuphorn, Parmesh Ramanathan and Kewal Saluja, *Sensor deployment strategy for target detection*, WSNA 2002

[43] Howard, A., Mataric´, M.J., and Sukhatme, A*n Incremental Self-Deployment Algorithm for Mobile Sensor Networks*, [Online] Available http://www.robotics.usc.edu/%7Eahoward/pubs/howard_ar01.pdf [November 2003]

[44] *Australian Climate*, [Online] Available http://www2.worldbook.com/educators/around_climate_australia.asp [September 2003]

[45] *Bufo marinus(Cane Toad)*, [Online] Available
http://animaldiversity.ummz.umich.edu/accounts/bufo/b._marinus$narrative.html,
[September 2003]

[46] *Pleb*, [Online] Available http://www.cse.unsw.edu.au/~pleb/, [August
2003]

[47] http://www.intellisense.com

[48] http://www.eyes.eu.org