Cane-toad Monitoring in Kakadu National Park Using Wireless Sensor Networks

Saurabh Shukla
School of Computer Science and Engineering
University of New South Wales
UNSW Sydney NSW 2052
Australia
sshu495@cse.unsw.edu.au

Nirupama Bulusu[*]
OGI School of Science and Engineering
Oregon Health and Sciences University
Beaverton, Oregon, USA
nbulusu@cse.ogi.edu

Sanjay Jha
School of Computer Science and Engineering
University of New South Wales and National ICT Australia
UNSW Sydney NSW 2052
Australia
sjha@cse.unsw.edu.au

ABSTRACT

This paper considers the problem of monitoring cane toads in Kakadu National Park using a large scale wireless sensor network deployment. Cane toads were mistakenly introduced in Australia 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. We comment on how wireless sensor network technology can address long-term research challenges for cane-toad monitoring, and propose a novel framework for planning sensor deployment to meet application, economic and networking objectives.

1. INTRODUCTION

This paper considers the problem of monitoring cane toads in Kakadu National Park using a large scale wireless sensor network deployment.

Cane-toad monitoring is an application of tremendous importance to Australia because cane toads are endangering native species and has attracted widespread national attention. The Kakadu National Park, a World Heritage Site 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 dark region in the Figure 1 shows the current distribution of cane toads in Australian Northern Territory. Traditional monitoring technology requires considerable human intervention and is inadequate to meet the long-term research objectives. Our objective is to deploy a large-scale network of inexpensive, lightweight sensors that are capable of acoustical observations to monitor, track and measure the impact of cane toads in Kakadu National Park. The sensor nodes receive analog data in form of toad vocalizations, digitize 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. We try to answer questions such as what kind of sensor devices to use, how many sensors to deploy, the density of sensors with in a region. We try to meet application objectives while minimizing the dollar costs of deployment.

[*]Work done while author was at National ICT Australia Limited, Bay 15, Locomotive Workshop Australian Technology Park Eveleigh, NSW 1430 Australia.

Figure 1: Current Distribution of cane toads in Australia.

In this paper, we provide a novel framework for studying deployment by integrating application, economic, and networking/technology objectives.

1. *Zone Division and Classification:* Division of deployment area into zones and classification of the zones based on deployment priorities .

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

3. Observation that it is hard to get initial deployment right due to uncertainty. Adaptive learning algorithm to reconfigure sensor deployment using Bayesian inference.

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

2. CANE TOAD MONITORING

In this section, we discuss state of the art research in cane toad monitoring and also review the long term research goals that motivate the use of wireless sensor networks technology.

2.1 Significance

The cane toads possess a remarkable ability to adapt to a wide range of habitats. [6]. They 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. A study conducted in 1990 by the Commonwealth Scientific and Industrial Research Organization showed that these toads 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 with ample water resources and abundant food supply. Funded by Australian Government Commonwealth researchers have initiated research applying modern gene technology to somehow deter the toads from marching all over the country. [6].

2.2 State-of-the-Art

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. Sixteen independent PLEB devices record rainfall, temperature and distinctive vocalizations or frog calls during the wet seasons when 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 [4].

Each PLEB costs 1000 Australian dollars and it will be very expensive to deploy them all over Kakadu. It can only detect the presence of cane toads but wont be able to infer the direction of movement of cane toads as there is no coordination among PLEB devices.

A gene that will stop cane-toad growth (CSIRO): At the Commonwealth Scientific and Industrial Research Organization (CSIRO) researchers 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.

2.3 Long-term Research Goals

Traditional monitoring techniques require considerable human intervention, which is not desirable. Some species might react adversely to human presence. Several 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.

Due to their large size and costs, the PLEBs (used for target vocalization) will be sparsely deployed. This would create gaps in the information received requiring the biologists to make generalizations. Long term research must address the following goals:

- Track the direction of migration of cane toads.
- Growth and 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.

A large scale deployment of sensors in the Kakadu National Park as discussed next can provide information that was previously impossible to obtain using traditional methods.

3. LARGE-SCALE SENSOR DEPLOYMENT

Our objective is to deploy a network 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.

We hope to equip biologists and ecologists with a tool to track and monitor the growth and impact of cane toad populations. The fully deployed system will be capable of pin pointing the count and concentration of cane toads in the Kakadu national park.

3.1 Design Goals

The ideal sensor network should support sensing coverage of a wide area of the Kakadu National Park , energy efficient, and robust to networking, data transmission and sensing errors. Moreover, the system should be able to process data efficiently. The network load is expected to be quite high during peak times and close to zero rest of the time. The system should programmed to optimize the trade off between data processing and transmission.

The ideal system would have the following functional characteristics:

1. *Data Storage*: 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 during the day.
2. *Detection Accuracy*: The it impossible for the system to learn from cane toad vocalizations 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.

3.2 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.

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

2. The overall spread of cane toads between catchments is 27 km/year.

3. Toads need to visit water every three days.

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

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

6. Cane toads can disperse by natural and by transport mediums used by humans. Roads and vehicle tracks provide traversal routes and also have open level ground along which the toads concentrate their activity. Faster natural means include transport of eggs and tadpoles by flowing water, swimming by adults in flood water and hiking.

3.3 Solution Approach

The proposed solution to sensor deployment consists of 4 steps.

1. *Zone Division and Categorization*: Division of deployment area into zones. Classification of zones based on deployment priorities.

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

3. *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. ZONE DIVISION AND CATEGORIZATION

The area of Kakadu National Park is around 200,000 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.

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 6.1) which comes after the first deployment. Due to the lack of metrics available to further categorize these zones, categorizations has been made broad.

An analysis of all 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.

5. IN-ZONE DEPLOYMENT

This section provides a description of the wireless sensor network deployment within a zone. The two requirements of our habitat monitoring system are target detection and target localization.

5.1 Target Vocalization

Target vocalization determines whether observed animal calls belong to cane toads using their spectrograms (species detection). 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 [4].

5.2 Target Localization

Target localization 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.

5.3 Components and Deployment

All nodes have integrated sensing, processing, and communication capabilities. 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 vocalizations expensive and time consuming. Grigg et. al. performed similar experiments and it was found that 15 seconds of sound takes 1 minute of process. The toads are active in the night and they tend to repeat their vocalizations 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 chose an approach where the deployed network consists of:

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)

Both micronodes and macronodes will have acoustic sensors and communication will be over a wireless 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 higher 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 tiered clustered wireless network(Figure 2) .

5.4 Deployment Strategy

Sensor nodes will typically be deployed in dense sensor patches that are widely separated. 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 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. The number of macronodes or cluster heads will be limited, bounded by cost constraints and bandwidth requirements.

5.5 Event Dynamics

Cane toad species are most active at night. 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. They will periodically go to sleep if no cane toad activity is detected.

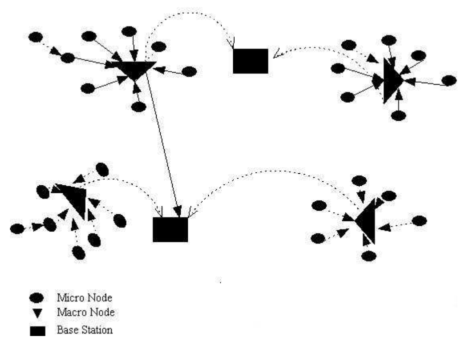

Figure 2: Sensor Network

Table 1: Spatial Density

Detection Range	*Area Covered*	*Spatial Density*
20 m	1040 sq m	.001 nodes/sq m

5.5.1 Spatial Density

The sensing fidelity generally depends on distance of the source from the microsensor node. The specific sensing function parameters depend on the nature of the sensor device and usually have the form d^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:

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

where $d(s,p)^k$ = the euclidean distance between the sensor node s and the cane toad at point p. γ 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 1 gives the minimum node density and the area covered by one node (with $k = 2$) to be able to detect toads in a zone.

5.5.2 Vocalization Processing

Cane toad vocalizations detected by the micronodes perform pre-processing of the vocalization to determine if it belongs to the cane toads. The vocalization received is sampled and a spectrogram is created. The decision that the vocalization matches the species to be observed is determined by the maximum cross-correlation coefficient between the observed spectrogram and the specified characteristic spectrogram which is input to the system. Figure 3 (a) shows the

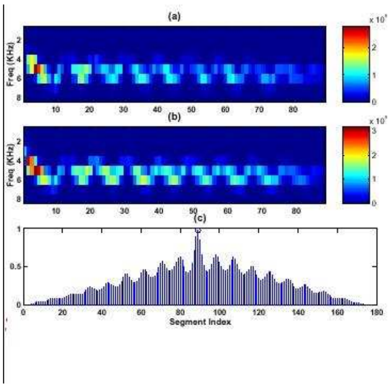

Figure 3: Spectrum of Frog

Table 2: Bandwidth Requirements

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

spectrum of observed vocalization, Figure 3 (b) shows the spectrum that is input to the system and Figure 3 (c) shows their cross-correlation coefficients. (source: [4])

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 matters is the one that in the same bandwidth as the animal calls. Table 5.5.2 shows the bandwidth requirements to process a cane toad vocalization.

The whole data processing task is divided into three stages: 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[3].

5.5.3 Target 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 iterative triangulation. To locate the calling animal when its call is recognized, the system proposed by Wang et. al. [4] 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 [5].

5.5.4 Data Compression

Even though micronode processing avoids transmitting raw data to cluster heard by processing data locally, the beamforming nodes still need the 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 to lower the transmitted data volume [3].

6. ADAPTIVE LEARNING

It is impossible to get the deployment right the first time. We use Bayesian networks to learn from previous deployments and adapt the next deployment.

6.1 Bayesian Belief Network

A *Bayesian network* is a graphical model that encodes probabilistic relationships among variables of interest[31]. A Bayesian network can be used to learn causal relationships, and hence can be used to gain understanding about a problem domain, even when all data entries are not known. 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. The Bayesian rule helps to estimate the most probable underlying model for a random process, based on some observed data or evidence.

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,

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)\, P(A_i) \tag{2}$$

{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:

$$P(A_k \mid B) = \frac{P(B \mid A_k)\, P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i) P(A_i)} \tag{3}$$

6.2 Application Variables

The following is a list of mutually exclusive independent events that influence the densities of Cane toads within a region. We base our model on three factors. (i) access to water, (ii) food resources, and (iii) access to the region.

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

1. *Access to Water.* Measures effect of water resources. Toads need to visit water once in three days. The toad tries utmost to keep moist and preserve body water levels. The cane toads need water to lay their eggs. We therefore categorize the effects of water resources into three categories depending on how far water resources are:

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

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

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

   (a) *Medium range*(w_2): Refers to the water bodies which are within a range(r_1) where 1 ¡ r_1 ¡ 3 km .

$$w_2 = \begin{cases} 0, & no\,water\,resources\,in\,range\,r_1 \\ 1, & water\,resources\,in\,range\,r_1 \end{cases}$$

   (a) *Far range(w_3):* Refers to the water bodies which are in range(r_2) where 3 ¡ r_2 ¡ 10 km.

$$w_3 = \begin{cases} 0, & no\,water\,resources\,in\,range\,r_2 \\ 1, & water\,resource\,in\,range\,r_2. \end{cases}$$

   Considering the speed at which toad moves, 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. 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. In most situations, the denser the vegetation, the greater the insect population it can support and hence more food is available for the cane toads. 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 (a.f.r)
     $$f_4 = \begin{cases} 0, & when\,a.f.r\,are\,not\,present \\ 1, & when\,a.f.r\,are\,present \end{cases}$$

     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. [6] found that the spread of cane toads in the Kakadu National Park will depend on the transport corridors available to them.

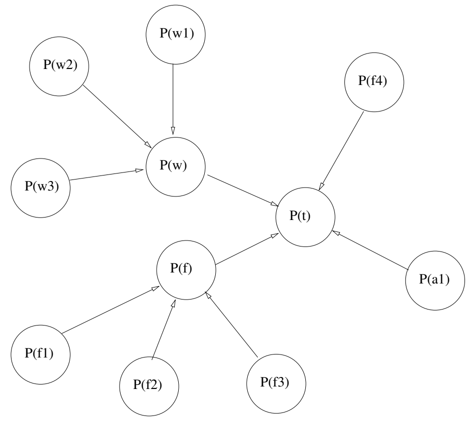

Figure 4: Bayes Network

   This includes pathways, roads, riverbeds and "hitch-hiking" in human vehicles.

   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'.

$$a_{i,j} = \begin{cases} 0, & when\,transport\,is\,unavailable \\ & from\,zone\,j\,to\,zone\,i \\ 1, & when\,transport\,is\,available \\ & from\,zone\,j\,to\,zone\,i \end{cases}$$

6.3 Intuitive Probabilities

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. 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 4 shows the variables in our Bayesian network and the Bayes Net that was built.

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, [1]

[1] 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.

Table 3: Distribution Table

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

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

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 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 3 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.

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 4) 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 consisting of finite number of such locations.

6.4 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

$$P(T \mid Z_j) = \frac{\sum_{i=1}^{32} q_i P(X_t = X_i)}{\sum_{i=1}^{32} q_i} \tag{4}$$

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

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

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

$P_j^u(attr) = \sum_{i=1}^{i=n}(aP(attr)_i + P(a_{i,j})) + bP(t \mid Z_j)$
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.

where $P(a_{i,j})$ refers to the access routes available from zone 'i' to zone 'j'.

where -	o	is occupied zone
	u	is unoccupied zone
	attr	is attraction level
	n	is all neighbouring zones

Therefore

$$P_j^o(attr) = \frac{\sum_{i=1}^{32} q_i P(X_t = X_i)}{\sum_{i=1}^{32} q_i} \quad (5)$$

$$P_j^u(attr) = \sum_{i=1}^{i=n} (aP(attr)_i + P(a_i)) + bP(t \mid Z_j) \quad (6)$$

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.

6.5 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 mudholes 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(attr) 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.[10]

- 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): Premonsoon 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 $s_{z\,s}\ P(attr)$ as the columns where P(attr) is derived from equation (5) and (6).

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.

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

In the above matrix, 1, 2, 3, 4 .. z_n are are the zones. The matrix can be used to find the maximum P(attr) 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(attr). 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.

The geographical region of Kakadu National Park is completely explored (Figure 5), 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).

7. COST

One of the metrics that determines how well an application has performed is the cost incurred for deployment. We study the *Direct cost*, which refers to the expenses pertaining to sensor nodes, base stations, macronodes (PLEBs) [12], and batteries. The dominant factor in 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 limited channel capacity, each macronode cannot support more than a certain maximum number of sensors. The number of PLEBs required within a zone will be determined by the bandwidth required to transmit the acoustic data from all the sensor nodes.

Sensors send data to the macronodes over a multi-hop wireless network arranged in a hierarchical structure. For a hexagonal cell structure (Figure 6), we can calculate that the data rate required increases exponentially with the number of hops to the macronode. The data rate is proportional

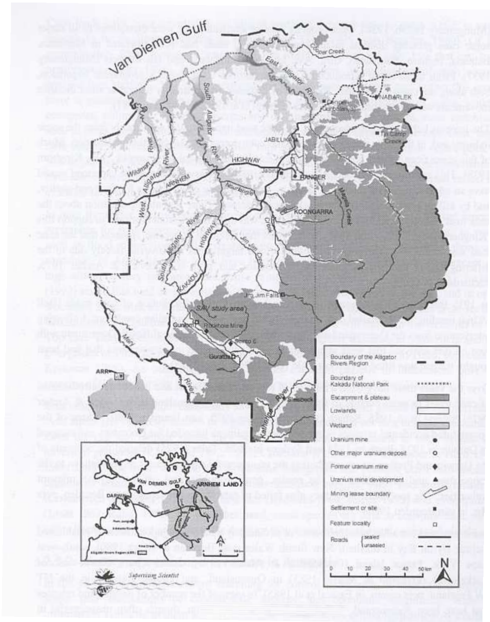

Figure 5: Map of Kakadu National Park

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, and 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 = $\frac{3\sqrt{3}d^2}{2} \approx 1040\ m^2$
Area of a Zone = $10^7 m^2$
Number of sensor nodes required =

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

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

Bandwidth Requirement: The highest vocalizations are of frequency 8 KHz. With 8 bit sampling, the data rate = 128 kbps.

Using compression algorithm with S-encoding compression of less than 1% can be acheived [9]. Therefore, data rate required at sensor node for one vocalization = 1.28 kbps

Assume the maximum data rate a sensor node can handle $\simeq$128 kbps. In the 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.

Number of sensor nodes under a PLEB =

$$\frac{Worst\,Case}{percent\,of\,active\,nodes} = 2000$$

Number of PLEBs in a zone =

$$\frac{Total\,Number\,of\,Nodes}{Number\,of\,Nodes\,per\,PLEB} = \frac{9600}{2000} \simeq 5$$

Cost Estimates: Let k be the number of PLEBs required. Assuming a sensor node costs c_1 and a PLEB costs c_2.

$$\text{Direct Cost} = n \times c_1 + k \times c_2$$

Table 4: Cost estimates for a zone (in US Dollars)

	Cost of Sensors	*Cost of* PLEBs	*Total Cost*
Only PLEBS	0	4.8×10^6	4.8×10^6
Only Sensor Nodes	2400	0	2400
Tiered architecture	2400	2500	4900

Table 5: Cost estimates for the Kakadu National Park (in US Dollars).

	Cost of Sensors	*Cost of* PLEBs	*Total Cost*
Only PLEBS	0	8680×10^6	8680×10^6
Only Sensor Nodes	4.32×10^6	0	4.32×10^6
Tiered Architecture	4.32×10^6	5×10^6	9.32×10^6

90% of Kakadu can be covered by sensor nodes, factoring the 90% we estimate the cost of instrumenting the entire Kakadu National Park. Tables 4 and 5 show the relative cost estimates for instrumenting one zone and the whole park with (a) only PLEBS, (b) only sensor nodes and (c) a combination of PLEBs and sensor nodes in a clustered environment respectively [2] . The final cost will be much lower

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

due to the hotspot identification by the adaptive algorithm proposed.

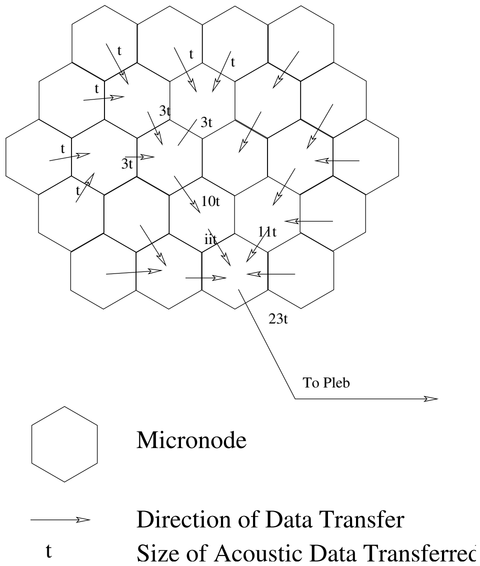

Figure 6: Data Transfer along micronodes.

8. CONCLUSIONS

In this paper, we considered network deployment and organization strategies for monitoring cane toads in Kakadu National Park using a distributed, tiered wireless sensor network. Although 90% of Kakadu is habitable by cane toads, they are expected to concentrate in a few regions which offer better food and ecosystems for their survival. The problem of sensor deployment was therefore reduced into *hotspot* zone classification (to meet economic objectives of deployment) and in-zone deployment of sensor nodes within hotspot zones (to meet sensing and networking objectives). To deal with uncertainties in the problem domain, an adaptive learning algorithm using Bayesian inference is proposed to identify and update the zones that might be future *hotspots* for cane toads to update deployment. By selectively identifying regions which are probable hotspots, we bring the deployment cost down.

9. ACKNOWLEDGMENTS

The authors would like to thank Dr. Andrew Taylor, UNSW for his general advice, Dr. Tony Robinson (CSIRO, Australia) for sharing his knowledge on cane toad behavior and the CSIRO research initiative on gene technology that stops their growth. Gernot Heiser and Adam Wiggins (UNSW and National ICT Australia) suggested future cost predictions of PLEB devices. Further, we wish to thank Matthias Hollick (TU Darmstadt) for discussions on modelling an infrastructure network and Dr R. W. Sutherst (CSIRO Entomology, Australia) for discussions on computer-modelling techniques to predict the geographical regions to be colonized by the cane toads.

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

[3] H. Wang, J. Elson, L. Girod, D. Estrin, and K. Yao. *Target classification and localization in habitat monitoring.* In ICASSP, volume 1, 2003.

[4] Taylor, A., G. Watson, G. Grigg, and H. McCallum. 1996. *Monitoring frog communities: an application of machine learning* AAAI/IAAI 2: 1564-1569. [8th Innovative Applications of Artificial Intelligence Conference, Portland, Oregon].

[5] K. Yao, R.E. Hudson, C.W. Reed, D. Chen, and F. Lorenzelli. *Blind beamforming on a randomly distributed sensor array system.* IEEE Journal of Selected Areas in Communications, 16(8):1555–1567, Oct 1998.

[6] 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.

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

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

[9] Hanbiao Wang, Deborah Estrin, Lewis Girod, *Preprocessing in a Tiered Sensor Network for Habitat Monitoring* EURASIP JASP special issue of sensor networks, Vol. 2003, No. 4, pp. 392-401, March 15, 2003.

[10] *Australian Climate*, [Online] Available `http://www2.worldbook.com/educators/around/climate/australia.asp` [September 2003]

[11] *Bufo marinus(Cane Toad)*, [Online] Available `http://animaldiversity.ummz.umich.edu/accounts/bufo/b_marinus/narrative.html`,[September 2003]

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