How to Do Research

How to writing High-Quality papers

How to Do Research

Doing good research =

Writing high quality papers to disseminate innovative solutions to interesting & challenging problems

Find a solution to a problem

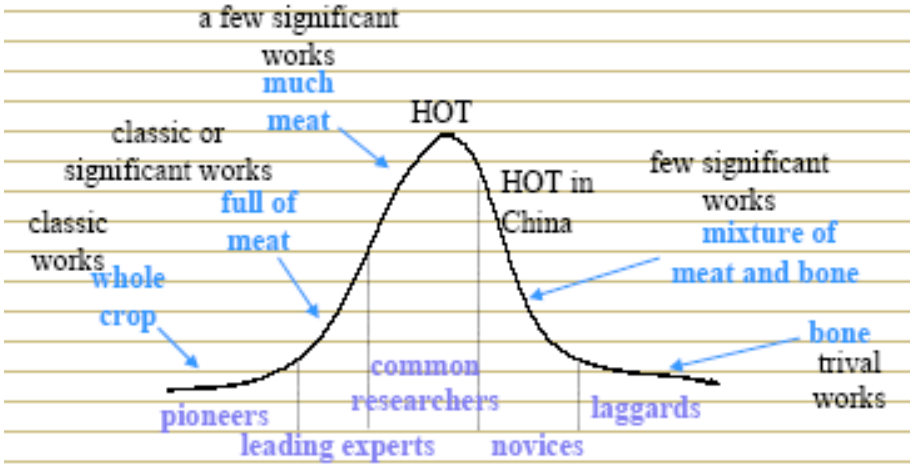
Not

Find a problem for a solution

Find a problem to solve

- **Problem must be of wide interest**
- **Problem must be significant & challenging**
  - Not toy, not trivial
  - Not only coding
- **Read papers in top journal / conference proceedings**
- **Find problems from applications**
- **Find problems by analyzing existing solutions to the problem**
  - Unsolved in some contexts

Find a problem to solve



Find a solution

- **Solution should be novel:**
  - A new method
  - An improvement to some existing method
  - A new way to use an old method
  - Combination of several methods
- **Know potential techniques**
- **Transform a new problem to some solvable problems**
- **Prove necessary & sufficient solutions**
- **Deep investigation + Sound theoretic analysis and well-covered experiments – solid (not superficial) results!**
  - Not only framework, not only model!

Keep the problem in mind all the time

- Know what exactly the problem is
- Feel excited about it
- Keep thinking

How?

- **Meetings and discussions**
  - Your supervisor
  - Your fellow students (in and out of the same group)
  - Be innovative
  - Make others to understand you!
- **Attend conferences, seminars, …**
- **Read papers and books, read a lot!**
  - Surveys & tutorials (classification, abstraction, future directions)
  - Journals, conference proceedings

The Hong Kong Polytechnic University 香港理工大學

How?

- **Know existing work and state of art**
- **Know leading groups and experts**
- **Know classified journals and conferences**
- **Know how to search**
- **Know how to throw away papers – read selectively**
- **Know how to read**
- **Know where to submit your papers**
  - Know acceptance rates
  - Know review criteria and possible recommendations

How?

- Does the paper introduce a new problem or provide a new solution to an existing one?
- What is the main result of this paper?
- Is the result significant?
- Is the paper technically sound?
- Does the paper provide an assessment of the strength and weakness of the results?
- Is the paper clearly written so as to accessible by most researchers in this area?
- Does the paper refer appropriate related works?

Writing High-Quality papers

What makes a good paper?

- **Contents**
  - Originality
  - Results
  - Contributions
- **Writing skills**
  - Organization
  - Presentation
  - Wording
- **Having high standard for both!**

Contents

- **Well-motivated, interesting problem**
  - With real-world applications or theoretical values
- **Challenging issues**
  - Worthy to be investigated
- **Good ideas and approach**
  - Technically sound?
- **Significant results**
- **Extensive & intensive analysis and evaluation**
- **Assessment of strengths and limitations**
- **Reference to appropriate related work**

"This is a very solid paper."

Contents

- **Originality, Novelty**
  - New problem?
    - Or, old problem / similar problem?
  - New solution / approach?
    - Or, just add / modify a bit?
  - New results?
    - Or, just minor improvement?

Contents

1. **Problem X is important**
2. **Previous work A, B have been studied**
3. **A,B have certain weakness**
4. **We propose our new method C**
5. **Experiment with C, compared with A,B**
6. **C is better than A,B (rigorously tested)**
7. **Why is C better? Why didn't D,E work?**
8. **Strength and weakness of C**
9. **Future work of C**

1-6 plus "organization of paper" => Introduction

Writing paper

General hints

- Enjoy the writing, not suffering – think of it as a learning experience !
- When writing the first draft (for you to read only), the goal is to put something down on paper, so it does not matter if sentences are incomplete and the grammar is incorrect, provided that the main points and ideas have been captured.
- Write quickly to keep the flow going.
- Write the paper in parts. Treat each section as a mini essay – think about the goal of that particular section and what you want to accomplish and say.
- Write when your energy is high , not when you are tired.
- Find a time and place where you can think and write without distractions.

Writing skills

General hints

- Put the first draft aside for at least one day, to allow you to "be" another person, switching between creation and critique.
- Know your reader!
  - In similar broad area, but may not have worked on your problem.
  - Keep in mind that you are writing for the reader, not for yourself.
- Keep revising the paper. Be prepared to do revision several times until you feel it is not possible to improve it further.
  - Revise for clarity and brevity.
  - Think in big, then think in small
  - Read it word by word.

Writing skills

General hints

- Not only write what you have done, but also why you do it.
- What you think in mind <> what you write in words
- Ask yourself questions, e.g.,
  - How do I express this?
  - Does each sentence make sense?
  - In your longer sentences, can you keep track of the subject at hand?
  - Do your longer paragraphs follow a single idea, or can they be broken into smaller paragraphs?

Writing paper

Organization and presentation

- Title of paper should be “eye catching”
- Organization of your paper should be logical – just like telling a story
  - Is it easy to follow?
  - Keep in mind all the time how the story should be developed - keep to the plan of your outline, don’t get lost.
  - Your English may be weak, but it in not an excuse for a weak paper, at least for structure and logic
  - Outline the structure before giving the details

Writing paper

Algorithm X has two components: A1 and B1. A1 is an improved version of …, which is, in turn, supported by A2. … We will describe X's first component in the order of A1, A2, and A3. Afterwards, we discuss B1 and B2 in details.

A1 …

A2 …

A3 …

Recall that X consists of A1 and B1, and we have already described A in details. We will now explain B in detail below.

B1 …

B2 …

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Writing paper

Organization and presentation

- Distinguish concepts and details, important and secondary contents; put them in appropriate order and places in the paper.
- When possible, formalize the presentation with definitions of concepts and terminologies, theoretical models, and analytical results.
- Number figures and tables separately using Arabic numbers. Make legends brief (no sentences!)

Writing paper

Organization and presentation

- Avoid parenthetical remarks. Avoid Run-on sentences (comma-separated)
  - If they are really important, elevate them to separate sentences or non-restrictive clauses. Or, use “but”, “and”, “because”
- Avoid putting two unrelated parts into one sentence
- Avoid numbered lists unless explicit reference is made to the items. Rewrite short unnumbered lists as paragraphs, and connect them logically.
- Avoid unnumbered headings.
- Avoid using bold and underlined phrases and sentences.
- Avoid having a paragraph with only one sentence.
- Avoid having a section with only one paragraph.

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Writing paper

Organization and presentation

- Don’t use the future tense unnecessarily.
- Use the active voice instead of passive voice.
- Use ‘which’ and ‘that’ correctly.
- Avoid using confusing sentences and words – clarify what you mean by short, firm, and clear sentences.
- Avoid using ‘this’ and ‘these’ alone – give the referent of ‘this’ and ‘these’ explicitly.

Writing paper

Organization and presentation

- Sell yourself – why & what are we good ?
  - Don't let the reader search for the interesting material – don't bury them in lengthy, useless paragraphs, but explicitly spell them out.
  - Pose potential questions and answer them yourself (e.g., what are your contributions?)
- Be concise and don't waste space for useless / repeating sentences and words. Edit your paper ruthlessly - check to see whether any sentence is meaningless or redundant; if so, delete it. 惜字如金！
  - For maximum readability, most sentences should be about 15-20 words.
  - Paragraphs of about 150 words in length are considered optimal.

Writing paper

Layout and Format

First of all, before you hand in a paper, make sure that your paper is well formatted.

- Double space, or 1.5 space the paper;
- Headings, fonts, font sizes, reference format are consistent
- Margins are well set and the paper should be justified on both sides
- Restrict headings to three levels (major and minor), and clearly distinguish between major and minor headings.

Paper template

- Abstract
- Introduction
- Review of Previous Work
- Our Work
- Experiments and Comparisons
- Conclusions
- Acknowledgement
- References
- Others (Appendix)

Writing paper

3. The Proposed Protocol

In this section, the details of the proposed consensus protocol are presented. We first introduce the system model and data structures, and then describe the operations of the protocol. The RZD property and the Look-Ahead technique of our protocol are discussed finally.

3.1 System Model and Data Structures

The system model for the proposed protocol is the same as in [1][5][11][19]. A distributed system consists of a finite set of n processes: $\Pi = \{p_1, p_2, \ldots, p_n\}$, $n > 1$. Processes communicate only by sending and receiving messages. Every pair of processes is connected by a reliable channel that does

IV. Fuzzy Control Model

A. Overview

In the Task Control Model, dynamic properties of the adaptation process are addressed at the end system, such as stability guarantees, adaptation agility and equilibrium fairness. However, the overall application behavior is non-linear and it may be possible that some desired QoS parameters cannot be maintained by simple parameter-tuning options. This section introduces the *Fuzzy Control Model*, which focuses on application-specific adaptation choices, with enhanced parameter-tuning possibilities or reconfigurations. The model utilizes results from fuzzy logic and the fuzzy control theory [10].

Paper template

Abstract

- Abstract is a VERY important part of your paper
- Purpose: Summary of your work and contributions
- Basically, here is your advertisement of your paper:
  - what you want to sell?
  - People decide to continue to read or now based on it
- Style?
  - What is the problem?
  - What is your solution and results?

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Paper template

Abstract

- Try to be concise but include the following
  - Describe the problem / issue addressed in the paper (What is it? Is it really important? Challenging?)
  - State the motivations (filling a gap? overcoming inadequateness and weakness of existing solutions?)
  - Highlight the ideas and features of your work (in comparison with existing ones – advantages? improvement?)
  - Highlight the main results of evaluation (analytical or experimental)

Paper template

Abstract

Naive Bayes is an efficient and effective learning algorithm, but previous results show that its representation ability is severely limited since it can only represent certain linearly separable functions in the binary domain. We give necessary and sufficient conditions on linearly separable functions in the binary domain to be learnable by Naive Bayes under uniform representation. We then show that the learnability (and error rates) of Naive Bayes can be affected dramatically by sampling distributions. Further, we demonstrate, through a specific example, that Naive Bayes can in fact represent non-linearly separable functions in the nominal domain. Our results help us to gain a much deeper understanding of this seemingly simple, yet powerful learning algorithm.

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Paper template

On Achieving Maximum Multicast Throughput in Undirected Networks

Abstract— The transmission of information within a data network is constrained by the network topology and link capacities. In this paper, we study the fundamental upper bound of information dissemination rates with these constraints in undirected networks, given the unique replicable and encodable properties of information flows. Based on recent advances in network coding and classical modelling techniques in flow networks, we provide a natural linear programming formulation of the maximum multicast rate problem. By applying Lagrangian relaxation on the primal and the dual LPs respectively, we derive (a) a necessary and sufficient condition characterizing multicast rate feasibility, and (b) an efficient and distributed subgradient algorithm for computing the maximum multicast rate. We also extend our discussions to multiple communication sessions, as well as to overlay and ad hoc network models. Both our theoretical and simulation results conclude that, network coding may not be instrumental to achieve better maximum multicast rates in most cases; rather, it facilitates the design of significantly more efficient algorithms to achieve such optimality.

Paper template

EnviroSuite: An Environmentally Immersive Programming Framework for Sensor Networks

Sensor networks open a new frontier for embedded distributed computing. Paradigms for sensor network programming in the large have been identified as a significant challenge towards developing large-scale applications. Classical programming languages are too low-level. This paper presents the design, implementation, and evaluation of *EnviroSuite*, a programming framework that introduces a new paradigm, called *environmentally immersive programming*, to abstract distributed interactions with the environment. Environmentally immersive programming refers to an object-based programming model in which individual objects represent physical elements in the external environment. It allows the programmer to think directly in terms of environmental abstractions. EnviroSuite provides language primitives for environmentally immersive programming that map transparently into a support library of distributed algorithms for tracking and environmental monitoring. We show how nesC code of realistic applications is significantly simplified using EnviroSuite, and demonstrate the resulting system performance on Mica2 and XSM platforms.

The Hong Kong Polytechnic University 香港理工大學

Paper template

Mobile Computing and Databases—A Survey

Abstract—The emergence of powerful portable computers, along with advances in wireless communication technologies, has made mobile computing a reality. Among the applications that are finding their way to the market of mobile computing—those that involve data management—hold a prominent position. In the past few years, there has been a tremendous surge of research in the area of data management in mobile computing. This research has produced interesting results in areas such as data dissemination over limited bandwith channels, location-dependent querying of data, and advanced interfaces for mobile computers. This paper is an effort to survey these techniques and to classify this research in a few broad areas.

Paper template

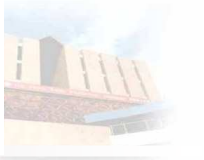
Introduction

- **One of the most important parts of your paper**
  - People usually read carefully on Abstract and Introduction to find out what is in your paper so as to decide whether it worth further reading, so don't make them disappointed
- **You can write Introduction by expanding the abstract**
  - Most important is to show how your work is motivated (background), focus of the paper, main ideas, and significance of results.
  - Sell your work - use *strong* tones!
- **Add outline of paper**

Paper template

Previous researchers (*e.g.*, (Salton *et al.* 1983; 1985)) have also considered the problem of relevance feedback for boolean query engines. Here an initial query (formulated by a user) is submitted to a search engine, and then modified based on labels collected from the user, after which the cycle is repeated. One difference in this paper is that we begin with a relatively large set of positive examples, rather than an initial query. A second difference is that we make use of an automated procedure to collect additional examples without user intervention, thus, making an assumption about the completeness of the current set of examples. This assumption is reasonable in this context, but not in a general information retrieval setting. The learning algorithms used in this paper are also novel in a relevance feedback setting.

Another point of difference with earlier work in information retrieval is that our experimental results are on



Paper template

I. Introduction

We study in this paper information dissemination in an undirected network, which consists of a set of end hosts and switches interconnected via undirected (or duplex) communication links. In data networks with known topologies and bandwidth capacity bounds for each undirected link, a fundamental problem is to compute and achieve the maximum end-to-end throughput for one or multiple active communication sessions. Depending on the objectives of applications, a communication session may be in the form of unicast (one-to-one), multicast (one-to-many), broadcast (one-to-all), or group communication (many-to-many). Our focus is on multicast, which is representative in that the other types of transmissions are special cases of or can be transformed into multicast transmissions.

Packet transmission in data networks may be modelled as the flow of bit streams, referred to as *information flows*. Compared to classical network flows, information flows may not only be buffered and forwarded, but also be replicated and coded. In previous work, it has been shown that by coding

Paper template

In this paper, we seek to bring new insights and efficient solutions to the problem of maximizing information flow rates (or *throughput*) in undirected data networks. We first illustrate the power of *network coding* with respect to achieving maximum throughput. Although previous directions of computing the maximum multicast rates involve solving NP-Complete problems, the maximum multicast rates and the corresponding optimal multicast strategy can indeed be computed efficiently in polynomial time, with the unique encodable property of information flows considered. We provide a natural linear programming formulation of the maximum throughput problem, with a polynomial number of variables and constraints. By applying Lagrangian relaxation on the primal LP, we derive a necessary and sufficient condition for multicast rate feasibility in undirected networks, from a distance labelling perspective. We show how it generalizes correspondent results in the unicast and broadcast cases, and how it connects multicast throughput with network capacity and bandwidth consumption. We further apply Lagrangian relaxation on the dual LP, and construct an efficient subgradient algorithm for computing the maximum multicast throughput and the corresponding optimal transmission scheme. We provide intuitive interpretations of the algorithm, and show that it can be implemented in a distributed fashion.

In addition, we extend the solution to multiple concurrent

Paper template

In addition, we extend the solution to multiple concurrent sessions without inter-session coding, as well as to other types of communication, including unicast, broadcast and group communication. Even when the general form of data networks is modified to reflect realistic characteristics of overlay networks (where only end hosts at the edge may be able to replicate and code data), or wireless ad hoc networks (where data is communicated through antennas), similar modelling and solution techniques are still effective.

Paper template

in realistically sized networks. We present empirical studies based on simulation results over thousands of test scenarios using our algorithms. We compare the maximum multicast rates with and without network coding, and show that noticeable gains can only be experienced in contrived network topologies; for random and irregular network topologies, such gain is almost always non-existent. This agrees with our theoretical results on the upper bound of the advantage of network coding

Our empirical studies also show that overlay multicast, which has recently attracted extensive research efforts, may be used to approach maximum rates quite well. To the best of our knowledge, this work is the first that systematically studies the effects of network coding with respect to maximizing information flow rates in *undirected* networks.

Paper template

1. INTRODUCTION

Sensor networks have been proposed for various applications including search and rescue, disaster relief, target tracking, and smart environments. The inherent characteristics of these sensor networks make a node's location an important part of their state. For such networks, location is being used to identify the location at which sensor readings originate, (for example, identifying a target's position during tracking, providing the location of an earthquake survivor buried underneath rubble). It is also used in communication protocols that route to geographical areas instead of IDs ([18][19][21][37]), and in other location based services, such as sensing coverage [38] and location directory service [22]. In addition to the applications and protocols mentioned, continued research in WSNs will serve to invent and identify many additional protocols and applications, many of which will likely depend on location aware sensing devices.

The Hong Kong Polytechnic University 香港理工大學

Paper template

Many localization algorithms for sensor networks have been proposed to provide per-node location information. With regard to the mechanisms used for estimating location, we divide these localization protocols into two categories: *range-based* and *range-free*. The former is defined by protocols that use absolute point-to-point distance estimates (range) or angle estimates for calculating location. The latter makes no assumption about the availability or validity of such information. Because of the hardware limitations of WSN devices, solutions in range-free localization are being pursued as a cost-effective alternative to more expensive range-based approaches.

Paper template

This paper makes three major contributions to the localization problem in WSNs. First, we propose a novel range-free algorithm, called APIT, with enhanced performance under realistic system configurations. Second, though many different protocols [4][24][28] have been proposed to solve the localization problem in a range-free context, no prior work has been done to compare them in realistic settings. This paper is the first to provide a realistic and detailed quantitative comparison of existing range-free algorithms to determine the system configurations under which each is optimized. We perform such a study to serve as a guide for future research. Third, no attempt has previously been made to broadly study the impact of location error on various location-dependent applications and protocols. This paper provides insight into the effect of localization accuracy on applications and suggestions on how to improve their performance in the presence of such inaccuracy.

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Paper template

The remainder of this paper is organized as follows. We first discuss related work in Sec. II. From Sec. III to Sec. VI, we present the feasibility condition and efficient solutions for the single multicast case. In Sec. VII, we extend our results to the cases of multiple sessions of unicast, multicast, broadcast, and group communication. We also consider the model of overlay networks and the model of wireless ad hoc networks. We then present empirical studies in Sec. VIII, and conclude the paper in Sec. IX.

Paper template

n of *index page*
pages that facil-
e.

stering problem
ering, which we
ing to partition
sters, we search
ossibly overlap-
ather, a cluster
r logs as input
ndex pages. Fi-
eGather is both
effective than
s task. Our ex-
l over a month

es

eadily yields its
lem of good web
rs. First, differ-
d, the same vis-
different times

sell and gather quite a bit of information from users. Providing such information to the site can be time-consuming and may be an invasion of privacy. *Optimization* is improving the site's structure based on interactions with all visitors. Instead of making changes for each individual, the site learns from numerous past visitors to make the site easier to use for all, including those who have never used it before.

While previous work has focused on customizing web sites, we chose to investigate web site optimization through the automatic synthesis of index pages. In the next section, we discuss our general approach and present the *index page synthesis problem*. We then present our technique, which we call cluster mining, and its instantiation in the PageGather algorithm; PageGather solves the subproblem of automatically synthesizing the set of links that comprises an index page. Following, we present the results of experiments run using our implemented system. Finally, we discuss related work and future directions.

The Index Page Synthesis Problem

The Hong Kong Polytechnic University 香港理工大學

Paper template

- **Related work**
  - Tell people the background and existing works related to your research
    - It can also be placed at the end of the paper (before Conclusion), depending on whether your work heavily depends on these works or not
  - Purpose: draw the differences

Paper template

- **Related work**
  - **Be lucid – summarize and classify existing works by describing the main approaches used and results of important works.**
    - Don't simply give a paper-by-paper description without logical development.
  - **Be critical but skillful in pointing out the weakness of existing works – don't overly criticize**
  - **Whenever possible, compare them with your proposed solution**
    - You can borrow ideas / techniques from them but must have some new stuff (improvement or extension, or apply them to new environment

Paper template

- **Related work**
  - **Style**
    - Previous work: may split to several classes
    - Can review each work in one or several sentences
    - Compare to yours (refer to later sections)
      - Emphasize the differences
      - Don't misinterpret
      - May also put it after sections about your own work

Paper template

As IP multicast is not readily deployed, algorithms promoting application-layer overlay multicast have recently been proposed as remedial solutions, focusing on the issue of constructing and maintaining a multicast tree using only end hosts [5], [6]. Though a single multicast tree may not lead to optimal throughput, recent studies (*e.g.,* SplitStream [7], CoopNet [8], Digital Fountain [9] and Bullet [10]) have proposed to utilize either multiple multicast trees (*forest*) or a topological *mesh* to deliver striped data from the source, using either multiple description coding or source erasure codes to split content to be multicast. These proposals have indeed improved end-to-end throughput beyond that of a single tree, but there have been no discussions on whether the optimal throughput may be achieved, or how close the proposed algorithms appr optimality. In this paper, we study such achievable optim while considering the general case where the data s transmits a stream of bytes, and is not assumed to per any source or error correction coding.

Traditional network flow theory studies the transmission of goods within a capacitied transportation network. The maximum transmission rate between two nodes is characterized by the celebrated max-flow min-cut theorem [14]: *a flow rate χ between nodes u and v is feasible, if and only if every cut between u and v has size at least χ*. Various algorithms may compute the maximum flow efficiently, some of which allow fully distributed implementation, *e.g.*, the push-relabel algorithm [14] and the ϵ-relaxation algorithm [15]. While information flows also need to confine to network topology and respect link capacities, they are different than commodity flows in that they are replicable and encodable. Data replication and

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Paper template

existing studies on cache invalidation and replacement strategies for mobile clients. Most of them were designed for general data services and only a few addressed the caching issues for location-dependent data.

As categorized in [22], [23], there are two kinds of cache invalidation methods for mobile databases: *temporal-dependent invalidation* and *location-dependent invalidation*. Temporal-dependent invalidation is caused by data updates. To carry out temporal-dependent invalidation, the server keeps track of the update history (for a reasonable length of time) and sends it, in the form of an *invalidation report* (*IR*), to the clients, either by periodic/aperiodic broadcasting or upon individual requests from the clients [2], [4], [11], [12]. In the basic *IR* approach, the server broadcasts a list of *ID*s for the items that have been changed within a history window. The mobile client, if active, listens to the *IR*s and updates its cache accordingly. Most of the e temporal-dependent invalidation schemes are variat the basic *IR* approach. They differ from one another in the organization of *IR* contents and the mechar uplink checking. A good survey can be found in [1

In location-dependent services, a previously cached data value may become invalid when the client moves to a new location. Location-dependent invalidation is due to mobile clients movements. In a previous paper [23], we assumed a cell-based symbolic location model and proposed three location-dependent invalidation schemes. Their performance was investigated using an analytical model in [22]. No location-dependent invalidation schemes have been proposed for a geometric location model.

Paper template

Many existing systems and protocols attempt to solve the problem of determining a node's location within its environment. The approaches taken to solve this localization problem differ in the assumptions that they make about their respective network and device capabilities. These include assumptions about device hardware, signal propagation models, timing and energy requirements, network makeup (homogeneous vs. heterogeneous), the nature of the environment (indoor vs. outdoor), node or beacon density, time synchronization of devices, communication costs, error requirements, and device mobility. In this section, we discuss prior work in localization with regard to these characteristics. We divide our discussion into two subsections where we present both range-based and range-free solutions.

2.1 Range-Based Localization Schemes

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Paper template

- **Our work**
  - **Purpose**: describe our work – may split to several sections
  - **Style**:
    - Definition, notation (need motivation) – in the shoes of the readers
    - Algorithms: pseudo-code; diagram; explanations
    - Answer potential questions from readers
    - Too much detail (e.g., proof): appendix
    - Exceptions: footnotes.

Paper template

Preliminaries

- Present your system architecture, system models, assumptions, concepts, notations, definitions, and some claiming theorems that are used in the rest of your paper
- You can also give inspiring examples / cases / scenarios
- Be accurate, as they are the basis of your work. If some of these change, your work or its quality may be affected.

Paper template

III. Task Control Model

A. Motivation

One of the major features of the *Task Control Model* that differs from previous adaptation schemes is that it focuses on *actively controlling* the adaptation behavior of applications, rather than simply providing hints to applications about current system states via upcalls, so that applications can adapt by itself. The justifications of adopting an *active control* approach are derived by the following observation of a typical control system in control theory [6].

If we examine the essential characteristics of the adaptation process, it corresponds naturally to a typical control system. In a control system, there is a target system to be controlled. The internal states within the target are determined by a *controller* according to a *control algorithm*. The output of the control algorithm is determined based on the states observed i such as individual mobility patterns. in actions to deliver a result to them to the desired val *setpoint*. The *observer* i

In this paper, *location granularity* is assumed to be a cell, which falls into the second category. The cell-based location model is well justified with the following two reasons. First, cell-based location identification requires neither additional devices deployed on mobile clients nor modifications over the current cellular network infrastructure. Thus, this is the cheapest solution. Second, with recent development in micro-cell/pico-cell systems,[3] it is believed that this model

the system, actually integrate the controller into the application, while leaving the system to serve as observers. We propose to detach the controller from the application, so that all concurrent applications can benefit from a *unified controller* design. The role of this controller is implemented by the *Adaptor* in the middleware control framework. This design strategy forms the basis of the Task Control Model.

The above design strategy adopted by the Task Control Model leads to two major advantages. First, because of the *unified controller* design, it is straightforward to utilize

B. Task Flow Model

In order to design control algorithms using the control theory, we need a strict mapping between models used in control systems and our design of active adaptation control. For this purpose, we consider each application as an ensemble of functional components, which we refer to as *tasks*. Tasks are execution units that consume system re-

on of *tasks*, we utilize the Task the modeling of active adapta- low Model, a Task Flow Graph,

Paper template

The send and receive events of a message m are denoted respectively with *send*(m) and *receive*(m). A *distributed execution* $\hat{E}$ can be modeled as a partial order of events $\hat{E} = (E, \rightarrow)$, where E is the set of all events and $\rightarrow$ is the *happened-before relation* [8] defined as follows:

DEFINITION 2.1. *An event $e_{i,h}$ precedes an event $e_{j,k}$, denoted $e_{i,h} \rightarrow e_{j,k}$, iff:*

- $i = j$ *and* $k = h + 1$, *or*
- $e_{i,h} = send(m)$ *and* $e_{j,k} = receive(m)$, *or*
- $\exists e_{l,z} : (e_{i,h} \rightarrow e_{l,z}) \wedge (e_{l,z} \rightarrow e_{j,k})$.

A checkpoint C dumps the current process sta stable storage. A checkpoint of process P_i is denoted

A *global checkpoint* C is a set of local checkpoints

$$\left\{C_{1,sn_1}, C_{2,sn_2}, \ldots, C_{n,sn_n}\right\}$$

one for each process.

DEFINITION 2.3. *A global checkpoint*

$$C = \left\{C_{1,sn_1}, C_{2,sn_2}, \ldots, C_{n,sn_n}\right\}$$

is consistent iff

$$\forall i, j \in [1, n] : i \neq j \Rightarrow \neg\left(C_{i,sn_i} \rightarrow_C C_{j,sn_j}\right).$$

In the following, we denote with C_{sn} a global checkpoin

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Paper template

First, let us define a new delay parameter

$$d = \alpha - \frac{N}{2} \tag{5}$$

and a sequence of numbers c_j as

$$c_j = \sum_{i=0}^{j} (-1)^{j-i} \binom{\frac{N}{2} - d}{i} \binom{\frac{N}{2} + d}{j - i}. \tag{6}$$

For noninteger values of d, the binomial coefficients involved in the above expression are evaluated using

$$\binom{x}{i} = \begin{cases} \prod_{j=0}^{i-1} \frac{x-j}{j+1}, & i \geq 1 \\ 1, & i = 0 \\ 0, & i < 0. \end{cases} \tag{7}$$

Now we assert the following.

Theorem 1: $c_j = b_j$ for all integers j.

cache consistency. To facilitate our discussion, the following notations are defined (note that these parameters are for one client only):

- D: the number of data items in the database.
- C: the size of the client cache.
- $\overline{a_i}$: mean access arrival rate of data item i,

$$i = 1, 2, \ldots, D.$$

Paper template

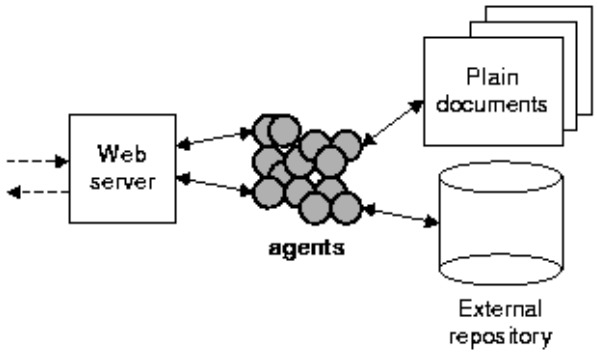


Figure 4: Agents receive a request through a Web server and control the documents.

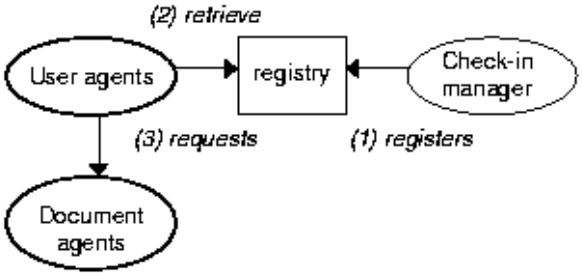


Figure 5: Key interactions among components in Persona.

Paper template

2 BACKGROUND

In this section, several potential applications for location-dependent information are first described to motivate our study. Afterwards, we provide some background on the location model, the mobile computing model and the temporal-dependent invalidation scheme adopted in this study.

2.1 Potential Applications

Two examples are described below to help us understand potential applications for location-dependent information. Other application scenarios could include community services, health, and entertainment etc., and more can be found on AltaVista Local.[2]

Example 1 (Travel Information). Suppose a traveler is visiting a new city and he wants to find a restaurant around lunch time, but he does not know his current

Writing paper

- **The main part of your work**
  - This part may consist of several subsections. Before the details, however, you may want to have a paragraph to briefly describe the contents and outline the organization of the subsections
  - **Data structures**
    - Describe the data structures and other building blocks, e.g., message types, that are used in algorithm/protocol

Writing paper

3.2 Data Structures and Message Types

When executing the protocol, each host m_i needs to maintain necessary information about its state. Such information is stored in the following variables:

- fl_i: The flag indicating whether m_i has made the decision.
- r_i: The sequence number of the current round in which m_i is participating.
- ph_i: The phase number of the current phase in which m_i is participating.
- est_i: The current estimate of the decision value. Initially, it is set to the value proposed by m_i.
- ts_i: The timestamp of est_i. The value is the seq number of the round in which m_i receives est_i the coordinator. The update of ts_i is entailed l reception of estimate from a coordinator.

The message types involved in the proposed protocol are listed as follows:

- $PROP(r, est_{cc})$: The proposal message sent from the coordinator to the clusterheads or from a clusterhead to the hosts in its cluster. est_{cc} is the current estimate kept by the coordinator. In each round, the coordinator tries to impose est_{cc} on other hosts by sending proposal messages.
- $ECHOL(r, est_i, ts_i)$: The echo message from m_i to its clusterhead in round r.
- $ECHOG(r, v, ts_v, x, y)$: The echo message from a clusterhead to other clusterheads in round r. It is

Writing paper

- **The main part of your work**
  - **Algorithm/Protocol**
    - Have an informal overview of your method
    - Give detailed description of major components and how they interact.
    - Use pseudocode and list it as a figure. If necessary, show line no. Give descriptions to the code!
    - Can be split into several parts, e.g., the basic algorithm, then handling of topology changes / node failures, etc.
    - Use figures and examples to help explain / illustrate operation, if necessary

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Writing paper

III. DDVMA: A HEURISTIC MULTICAST ALGORITHM

A. Overview of DDVMA

DDVMA constructs a QoS multicast tree over the backbone network to transmit multicast messages from the source gateway to all the destination gateways. An optimal wireless route between each leader MH and its gateway is discovered by the AODV routing protocol. The delay values of the wireless routes are collected for computation in DDVMA.

Comparing to DDVCA, the improvement of DDVMA is realized by using the proprietary second shortest path or partially proprietary second shortest path to replace the multicast path with the minimum end-to-end delay on the SPT. The improvement procedure can be seen as an optimization procedure, i.e., using a better path to optimize the QoS of the SPT. The optimization objective is to achieve smaller multicast delay variation under multicast end-to-end delay constraint.

B. Formal Description of DDVMA

In this section, we will present a formal description of DDVMA as shown in Fig. 2. Two procedures are used, one is

Writing

Algorithm 1 Selection of the Best Valid Scope for the *CEB* Method

Input: valid scope $v = p(e_1, \cdots, e_n)$ of a data value;
Output: the attached valid scope v', $v' \subseteq v$;
Procedure:

1: v'_1 := the inscribed circle of $p(e_1, \cdots, e_n)$;
2: $v' := v'_1$; $E_{max} := E(v'_1)$;
3: $v'_i = p(e_1, \cdots, e_n)$;
4: $i := 2$;
5: **while** $n - i \geq 1$
6: //{containing at least three end-points for a polygon}
7: **if** $E(v'_i) > E_{max}$ **then**
8: $v' := v'_i$; $E_{max} := E(v'_i)$;
9: **end if**
10: **if** $n - i > 1$
11: $v'_{i+1} :=$ the polygon that is deleted one endpoint from v'_i while being bounded by v and has the maximal area;
12: **end if**
13: $i := i + 1$;
14: **end while**
15: output v'.

In our algorithm, each process keeps the identifiers of all its parents and children. Initially, variables $wait\_list(p_i)$ and $visitor(p_i)$ associated with each process p_i are empty. Assume that, currently, p_i requests a resource which is

```
procedure IREP(Pos,Neg)
begin
   Ruleset := ∅
   while Pos≠ ∅ do
      /* grow and prune a new rule */
      split (Pos,Neg) into (GrowPos,GrowNeg)
        and (PrunePos,PruneNeg)
      Rule := GrowRule(GrowPos,GrowNeg)
      Rule := PruneRule(Rule,PrunePos,PruneNeg)
      if the error rate of Rule on
        (PrunePos,PruneNeg) exceeds 50% then
          return Ruleset
      else
          add Rule to Ruleset
          remove examples covered by Rule
            from (Pos,Neg)
      endif
   endwhile
   return Ruleset
end
```

Figure 1: The IREP algorithm

Our algorithm is message driven. Each message is associated with a routine called *m-routine*. When a message is received, its m-routine will be executed. Before presenting our distributed algorithm, we first describe its main ideas. Assume that process p_i requests a resource which is currently held by p_j. The resource manager will generate a $WAIT(p_i, p_j)$ message and invoke our distributed algorithm. Consider the case when $WAIT(p_i, p_j)$ is received. If and $r(p_i)$ and $r(p_j)$ can be modified such that without breaching the dimensional limit of

Writing paper

D. An Illustrative Example of DDVMA

In the following, we will illustrate the operation of DDVMA with an example. We will contrast it with DDVCA, so we use the computer network topology given in [13]. The network topology is shown in Fig. 3. For a group communication scenario, we denote Vs as the source gateway, V2, V5 and V9

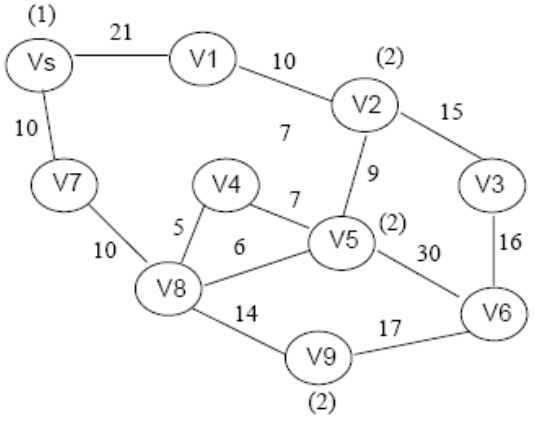
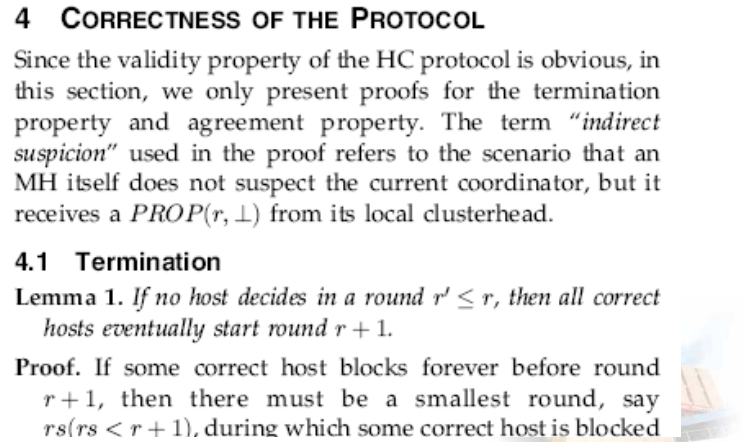


Fig. 3. A given network topology G=(V, E).

Writing paper

- **The main part of your work**
  - State and then prove / argue the properties of your solution, if necessary
    - e.g., correctness



4 CORRECTNESS OF THE PROTOCOL

Since the validity property of the HC protocol is obvious, in this section, we only present proofs for the termination property and agreement property. The term "*indirect suspicion*" used in the proof refers to the scenario that an MH itself does not suspect the current coordinator, but it receives a $PROP(r, \perp)$ from its local clusterhead.

4.1 Termination

Lemma 1. *If no host decides in a round $r' \leq r$, then all correct hosts eventually start round $r + 1$.*

Proof. If some correct host blocks forever before round $r + 1$, then there must be a smallest round, say $rs (rs < r + 1)$, during which some correct host is blocked

The Hong Kong Polytechnic University 香港理工大學

Writing paper

- **Performance analysis and / or Experiments**
  - A very important part of your paper, without this people will never be convinced of the quality of your work – and your paper will never be accepted!
    - Pay a great attention to this part, although it can be the most time-consuming part.
  - Clearly describe the models for theoretical analysis and experiments (simulations, real deployment)
  - Give details on
    - assumptions,
    - experiment / simulation setup,
    - parameter setting (are they reasonable?),
    - performance metrics, and

Writing paper

8. EVALUATION

In this evaluation section, we demonstrate the improved performance generated by our scheme in terms of 1) total amount of energy consumed, 2) energy variation among nodes, 3) half-life of the network and 4) sensing coverage over time. Moreover, we also demonstrate how we optimize the performance by considering the target size and unbalanced initial power that are not supported by previous schemes.

In the evaluation we do not include communication cost due to data transfer because it is highly application specific. Also for some applications as intruder detection, data transfer only happens when some rare events are issued. For such systems, energy spent on data transfer is relatively insignificant.

8.1 Simulation Configuration

We run our basic protocol and extensions on a special purpose simulator. In our simulation, the sensor nodes are distributed in a 160m × 160m square field. The sensing range is 10m and the communication range is 25m. The sensor nodes are deployed with a uniform distribution into the square field, unless otherwise stated. For our protocol, the target area is the 140m × 140m square in the center of the square field to prevent the nodes at the edge from working all the time. We only do statistics on the central 100m × 100m field to eliminate the edge effect. All experiments are repeated 10 times with different random seeds and different node deployments. The 95% confidence intervals of the results are about 5~10% of the means.

Writing paper

- **Performance analysis and / or simulation**
  - Illustrate the results in figures with curves, bar charts, etc. You can also use tables for comparison.
  - Don't forget to discuss the results – make sense of the results, don't just simply state the results!
    - describe your observations from the figures to provide insights into the result Have a summarizing paragraph at the end of this section – how good is your solution? When does it perform the best? Etc.
  - Comparisons (are they scientific? fair?)

Writing paper

8.2 Total Energy Conservation for the Basic Design

In this experiment, we investigate the energy conservation performance in term of total energy consumed per unit of time with the energy drain rate of one unit per round if the sensor node is awake all the time. We collect results from our basic design, the second pass optimization of our design and [Tian and Georganas 2003]'s sponsored coverage scheme. We also compare the simulation results with the lower bound and the upper bounds.

From Figure 11, we can see that our protocol consumes much less energy than [Tian

8.3 Balancing the Energy Consumption

In this simulation, we investigate performance of energy balance. We measure the standard deviation of energy consumed by each node in our basic design and in the multiple round extension with M=10. Results are compared with the sponsored coverage scheme.

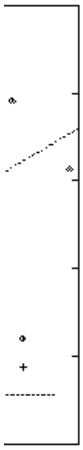
7.5 Performance vs. Non-sentry Duty cycle

Here, we evaluate the impact of the wake-up operation on the delay and energy consumption. First, the simulation results confirm that the average wake-up delay is approximately half of the toggle period as predicted in Section 4.3. Since the wake-up delay T_{wakeup} is one order of magnitude smaller than other delays such

Writing paper

We see from the Figure 13 that the distinguishing feature of our approach is that the system half-life increases nearly linearly as the node density increases, while the sponsored coverage scheme increases slowly when the node density increases. For example, our approach increases the half-life of the network by 130% when node density is 4 per r^2. There are two reasons contributing to this phenomenon. First, the sponsored coverage scheme consumes more energy on average than our approach. Second, the standard deviation among nodes in the sponsored coverage scheme increases significantly when the node density increases as shown in Figure 13. This causes some nodes dying faster than others.

Writing paper



axis) with
npling (y
and type

Several questions arise in connection with MetaCost's results. How sensitive are they to the number of resamples used? Would it be enough to simply use the class probabilities produced by a single run of the error-based classifier on the full training set? Would MetaCost perform better if all models were used in relabeling an example, irrespective of whether the example was used to learn them or not? And how well would MetaCost do if the class probabilities produced by C4.5R were ignored, and the probability of a class was estimated simply as the fraction of models that predicted it? This section answers these questions by carrying out the relevant experiments. For the sake of space, only results on the two-class databases are presented; the results on multiclass databases were broadly similar. Table 4 reports the results obtained for $r = 2$, 5 and 10 by the following variations of MetaCost: using 20 and 10 resamples instead of 50 (labeled "m=20" and "m=10"); relabeling the training examples using the class probabilities produced by

Writing paper

Appendix

Dataset	Naïve	TAN	HCS	SP	TAN Time	SP Time
Vehicle	61.97 ±1.58	63.47 ±2.47	**70.17** ±1.87	**70.25** ±2.01	333	1390
Post-op	70.01 ±0.83	70.06 ±1.53	**72.82** ±1.52	**72.13** ±2.11	+	+1
Lung	47.87 ±7.34	54.82 ±8.99	**58.34** ±8.18	**59.12** ±6.58	156	767
Australia	80.72 ±0.68	80.42 ±0.66	**84.74** ±0.75	**85.20** ±0.55	168	1299
Hepatitis	83.25 ±1.37	83.50 ±2.43	**84.75** ±1.87	**84.25** ±2.13	9	126
Vote	90.34 ±0.78	93.91 ±1.48	**95.58** ±0.56	**95.71** ±0.43	17	84
Heart	72.51 ±3.30	73.52 ±2.84	**78.73** ±2.16	**76.10** ±1.96	6	93
Soybean-Large	86.07 ±1.19	82.04 ±1.72	**88.83** ±1.22	**88.41** ±1.71	1046	13807
Pima	69.56 ±1.35	75.47 ±1.75	**78.00** ±1.31	**78.22** ±1.28	+	63
Breast	96.02 ±0.45	96.45 ±0.72	**97.41** ±0.89	96.12 ±0.81	21	172
Iris	**93.00** ±1.00	**93.60** ±0.95	**94.00** ±1.35	**93.60** ±1.25	3	10
Segment	90.92 ±1.86	86.25 ±1.65	**95.67** ±1.07	94.45 ±1.36	5491	62410
E.coli	80.21 ±0.44	80.89 ±0.69	**85.43** ±0.75	**84.35** ±0.34	16	91
exclusive-or	51.92 ±2.30	54.52 ±2.16	**68.22** ±1.46	**70.71** ±1.43	12	96

Table 6: Experimental results of comparing various algorithms. The best result and those not significantly worse than the best at the 5% confidence level are shown in bold. The last two columns contain the average time (in seconds) taken to build a classifier using TAN and SP

Writing paper

Conclusions

- Important – people often read Introduction and Conclusions first, so never overlook it.
- This section corresponds to the problem mentioned in Introduction – recite it.
- Then a quick summary of what you have written in the paper, mainly on your own achievements.
  - Tell people what you learnt from the study and what the most interesting issues are
- Future work -describe what can be improved and what you plan to do in your future work
- Closing- give some problems/issues remain open / challenging to solve, if any.

Writing paper

ɔm the one con-
ebert (Pazzani *et*
ith respect to a
milar to the one
al of this system
lists according to
e directory main-
ɛ Webert system
ıt retrieves addi-
ver, this (single)
simple heuristic
e an accurate in-

e work

rently filled with
collect together
cific topic. Keep-
ifficult due to the

of machine learn-
resource directo-

lists, while still attaining precision of greater than 90%.

The results reported in this paper suggest a number of topics for further research. The evaluations in this paper were of somewhat limited scope. Evaluation of the learning methods used in this study could also be conducted on artificial data, perhaps along the lines of the studies conducted in the Text Retrieval and Classification (TREC) meetings (Harman 1995). This would allow more rigorous comparative evaluation of learning methods, albeit in a somewhat more artificial setting.

Several future research goals involve improving the interaction between the learned rule and the various search engines. First, the addition of new search engines would be simplified by an automatic procedure to construct the transformation from a rule (produced by the learning subsystem) to a query (for the search engine). Second, since none of the search engines is clearly superior to the others, it may be better to use the results of all the available search engines rather than rely on a single engine; one possibility would be to use learning techniques to combine the output of different engines. Lastly, several search engines sup-

Writing paper

Acknowledgements

ACKNOWLEDGMENTS

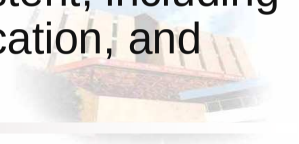
The authors would like to thank Tin-Fook Ngai, Qinglong Hu, and the anonymous reviewers for their valuable comments and suggestions that improved the quality of this paper. The research was supported by the Research Grant Council, Hong Kong SAR, China, under grant numbers HKUST-6077/97E and HKUST-6241/00E.

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Writing paper

References

- Purpose: supporting claims; know well all previous work
- People often check this list – sometimes for their own interest, but will criticize if disappointed.
- Give an adequate, up-to-date list of papers, they must be most representative – cite important, influential papers in well-known journals and conferences; also cite your own related papers
- Never overlook any important paper – people will think you are ignorant!
- Make sure the format of references is consistent, including authors, paper title, venue and time of publication, and pae numbers (if available)



Writing paper

- 写作是要花功夫，花心思的, 而且要细心
- 对文字要敏感！讲逻辑，将对称，讲一致
- 句子要简洁，不啰嗦，不重复(要强调时除外). 同时，意思要完整， 不要有二义性
- 用专业语言，标准格式（题目，布局，间距，图表，等等）
- 不要用长句子, 不要用短短落, 更不要一个Section只有一个段落
- 写完后，要打印出来，从头至尾，仔细阅读. 读不顺口的时候，往往有问题，不要放过