Communication in the Presence of Noise & Algorithms for Error-Correction MIT2001-04

Progress Report: July 1, 2001—December 31, 2001

Madhu Sudan

Project Overview

The project investigates fundamental problems arising from the theory of communication of information, from the perspective of algorithms, and computational complexity. The project centers on the theme of "list-decoding'', a notion that allows decoding algorithms to output a list of codewords that are close to a word received from a noisy channel. List-decoding allows for more errors to be recovered in a worst-case scenario, and offers a promising route to bridge the "probabilistic analysis", central to Shannon's theory of information, with more adversarial scenarios. The project investigates combinatorial, algorithmic, and complexity-theoretic aspects of list-decoding, including the task of finding very efficient algorithms for correcting errors in very noisy channels. The broader scope of this project includes research in algorithmic problems arising in all aspects of data communication.

Progress Through December 2001

Research has been carried in most of the directions listed in the project overview. The most significant developments in the central path (list-decoding) has been the development of very efficient algorithms for list-decoding. Along the broad scope, the main development is a worst-case analysis of "Grammar-based" compression algorithms. Other directions of progress include the study of the combinatorics of list-decoding, and applications of list-decoding to "guessing secrets". A summary of publications from this project includes:

- Conferences: 1 published, 6 accepted, 1 in preparation.
- Journals: 1 accepted, 5 in preparation.
- Theses: 1 Ph.D. theses filed, 1 under preparation.

In addition, the PI has given two invited talks in leading conferences (IEEE FOCS 2001, and AAECC 2001) on the research of this project, and coverage of his work includes an article in the SIAM News Magazine, January 2002. Below some of these achievements are described in detail.

Efficient algorithms for list-decoding:

Expander based codes: This result, obtained by Venkatesan Guruswami, under the supervision of Prof. Madhu Sudan, jointly with Prof. Piotr Indyk, shows that expander-based codes can be list-decoded efficiently upto a number of errors that matches the performance of best-known codes; but in only quadratic time. This result was presented at FOCS 2001.

Reed-Solomon codes: This result, obtained by Michael Aleknovich, under the supervision of Prof. Sudan, shows that the Guruswami-Sudan list-decoding algorithm can be speeded up to run in nearly-linear time, thus showing significant theoretical potential of list-decoding of Reed-Solomon codes. This result is being written up.

Grammar-based compression: The notion of compressing text is central to efficient transmission and storage of information. The idea of representing a string by a context-free grammar that generates it has been proposed in the literature. This result, obtained by Eric Lehman, April Rasala, and abhi shelat, under the supervision of Prof. Madhu Sudan, gives efficient algorithms to compute small grammars that produce a given string. The algorithms are accompanied by worst-case guarantees that are vastly better than prior results. The results have been accepted to SODA 2002, and STOC 2002. A full version is in preparation.

Guessing secrets: This result describes new applications of list-decoding to a task that arises in routing of traffic on the internet. The task is that of determining the IP addresses of clients by asking them a few questions about their address, where the clients may answer according to one of several IP addresses they may possess. The result shows how error-correcting codes and list-decoding algorithms can be used to design questions and recover the answers efficiently. This work is accepted to SODA 2002 and a full version is in preparation.

Other topics that were investigated include the combinatorics of list-decoding, unification of algebraic algorithms for list-decoding, and the harmonic broadcasting protocol.

Venkatesan Guruswami filed his Ph.D. thesis on the topic “List-decoding of error-correcting codes” in August 2001. Eric Lehman will be filing his Ph.D. thesis in February 2002 on “Grammar-based compression algorithms”.

Research Plan for the Next Six Months

The research plan for the next six months includes a continuation of ongoing investigation and initiation of some new topics.

Topics for continuing investigations include the design of efficient list-decoding algorithms, the study of the combinatorics of list-decoding, and the search for explicit codes that possess nice list-decoding properties. Abstracting existing list-decoding algorithms and classifying them will also be a component of the continued research.

New directions for research include:

- List-decoding and packing of information: The probability that a “bit” of information stored on a storage medium is reported incorrectly, is a function of the density with which one attempts to pack information on the storage medium. Classically, one attempts to minimize the probability of error on every bit since correcting too many errors was considered infeasible. However, list-decoding can now correct many more errors than was classical. This new ability calls for a reinvestigation of some of the classical choices on how to pack a storage medium. This project will research this question.

- When error-correcting codes are used to set up an interactive communication, then its structure is somewhat constrained. Codewords should be generated in an online fashion, and prefixes of codewords should form error-correcting codes. It is known that such codes exist, but constructive results are unknown. This project will investigate this question, as part of its general study of problems in information communication.
- List-decoding algorithms seems to be of use in the development of many fundamental primitives in cryptography. One such application, developed in the works of Boneh & Shaw, Staddon, Silverberg, and Walker, is the use of error-correcting codes to build watermarking, and fingerprinting schemes. Many aspects of these applications seem open to systematic investigation and this project will examine some of the fundamental parameters for optimality.