Research Statement **Date:** April 5, 2009
Ragesh Jaiswal[1]

1 Direct Product Theorems

My main research focuses on *Direct Product Theorems* and their applications in *Cryptography*, *Derandomization*, *Average-case Complexity*, and *Error Correcting Codes.* Direct Product Theorems are more formal statements with the following general intuition:

> *"if there is a problem which is hard to solve on the average, then solving multiple instances of the problem becomes even harder."*

Such theorems are useful in the following settings:

(i) Cryptography: Much of Cryptography is based on existence of problems which are hard to solve on average. Direct Product Theorems provide a consistent way to amplify security properties of certain cryptographic protocols. For instance, [IJK07] uses such theorems to show amplification of the gap between the success of a human user and a computer in solving a CAPTCHA[2] test in order to distinguish between the two.

(ii) Derandomization: A series of results (e.g. [NW94, IW97]) show a very interesting *Hardness-vs-Randomness* tradeoff. These results show the following sequence of implications: if there is a function which is hard in the worst-case, then there exists a function which is mildly hard-on-average. A Direct Product construction is then used to amplify the average-case hardness of such functions. The harder function is then used to construct a *pseudorandom* generator. This generator, which can extend a random seed by an exponential amount, is finally used to derandomize probabilistic computation. In short, a non-trivial circuit lower bound implies that randomness does not help as long as efficient computation is concerned.

(iii) Average-case Complexity: In the theory of average-case complexity, Direct Product Theorems help to address questions of the form: if there is a function within some worst-case complexity class $\mathcal{C}$ which is hard with respect to probabilistic computation, then is there another function in $\mathcal{C}$ which is even harder? More specifically, there is great interest in studying uniform hardness amplification within the complexity class NP (e.g. [O'D04, Tre05]).

(iv) Error Correcting Codes: Most of the proofs of Direct Product Theorems are constructive. This essentially means that the proof is by contradiction and has the following structure: starting with a solver which computes the strongly hard function on the average, the proof explicitly constructs a solver which computes the mildly hard function on the average. In certain cases, the construction of the strongly hard function can be viewed as an error correcting code, whereas the constructive proof of the theorem, gives a decoding algorithm. We will see such an error correcting code here.

Suppose we are given a boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$ which is hard on average for bounded size circuits. The classical Direct Product Theorem shows that the function f^k defined as $f^k(x_1, ..., x_k) = f(x_1)|...|f(x_k)$ ($|$ denotes concatenation) is much harder. Almost synonymous with the Direct Product Theorem is the classical Yao's XOR Lemma which says that the boolean function $f^{\oplus k}$ defined as $f^{\oplus k}(x_1, ..., x_k) = f(x_1) \oplus ... \oplus f(x_k)$ is much harder to compute than f. The proof of these Theorems is given by contradiction; that is, if the harder function f^k ($f^{\oplus k}$) can be computed by a circuit C (considering a nonuniform setting) of size s on at least ϵ fraction of the inputs, then

[1] Current affiliation: Postdoctoral Researcher at the Department of Computer Science, Columbia University. rjaiswal@cs.columbia.edu

[2] CAPTCHA was coined in [ABHL03] and is an abbreviation for Completely Automated Public Turing Test to tell Computers and Humans Apart.

we give an explicit construction of a circuit C' of certain smaller size $s' < s$ which computes f on at least $(1 - \delta)$ fraction of the inputs, where ϵ is exponentially small in k. Such constructions require random samples (input-output pairs) from f, which in the nonuniform setting is accounted for as the nonuniform *advice*. Next, I will discuss some specific instances of Direct Product Theorems and applications that I have been interested in.

1.1 Uniform Direct Product Theorems

Trevisan [T03] and Impagliazzo [I02] independently make the following interesting observation regarding the connection between the Direct Product Theorem and Error Correcting Codes: Considering the truth table of f as the message, the truth table of f^k can be interpreted as the k-wise Direct Product encoding of the message. The constructive proof of the Direct Product Theorem then gives an algorithm for list-decoding this code. Given this, we observe that the list size that all previously known proofs give is exponentially large ($2^{poly(1/\epsilon)}$) compared to the optimal (which should be $poly(1/\epsilon)$). This makes decoding using such constructive proofs highly inefficient. In joint works [IJK06, IJKW08] with Russell Impagliazzo, Valentine Kabanets and Avi Wigderson, we give a list decoding algorithm which is optimal with respect to the list size and running time. This also gives us an *advice-efficient* proof of the XOR Lemma which is used to show uniform hardness amplification within the the complexity class $\mathsf{P}^{\mathsf{NP}_{||}}$. Following are some interesting open questions of this work:

(i) This work considers the case when the inputs for f^k are chosen independently. It will be interesting to obtain an advice-efficient Direct Product Theorem even when the inputs are not completely independent. In other words, we want to obtain a "derandomized" version of our theorems. In [IJKW08] we obtain such a derandomized version. Further derandomization seems possible and is a subject of future investigation.

(ii) Our advice-efficient XOR Lemma helped us to show a uniform hardness amplification result within the complexity class $\mathsf{P}^{\mathsf{NP}_{||}}$. A bottleneck in showing uniform hardness amplification within NP is that even if f is in NP, $f^{\oplus k}$ is not necessarily in NP (unless $\mathsf{NP} = \mathsf{co} - \mathsf{NP}$). So the question is, can we show similar results for a monotone combination function instead of $\oplus$? That would give a uniform hardness amplification result within NP.

Our theorems have recently found an interesting application in a seemingly unrelated problem known as Proofs of Retrievability (PoR). The basic idea is to construct a scheme such that an untrusted server, which stores some client data, should be able to prove to the client that it holds its data with an efficient audit protocol. One of the popular PoR schemes [JK07] uses Direct Product construction and our theorems turn out to be very useful in showing efficiency of their scheme [DVW09].

1.2 Chernoff-type Direct Product Theorems

Classical Direct Product Theorems are statements of the following form: If a problem is hard on at least δ fraction of the instances, then answering k independent instances of the problem is exponentially harder to solve. Intuitively, the following statement also seems plausible: If a problem is hard on δ fraction of the instances, then making mistakes on smaller than δk problems from the list of k independent problems should drop exponentially. In a joint work [IJK07] with Russell Impagliazzo and Valentine Kabanets, we make this intuition precise by showing, what we call, a "Chernoff-Type" Direct Product Theorem. We prove this theorem for a very generic class of problems called *weakly verifiable puzzles* which has been studied in the Cryptography community [CHS05]. These puzzles capture two round challenge-response protocols (for example the CAPTCHA protocol).

This immediately finds the following application in Cryptography: Consider a two round challenge-response protocol where even a legitimate party has some chance of failure. For example, in the CAPTCHA test, even a human user can make a mistake sometimes due to a typing error or mis-reading. In this case, we want to amplify the security of the protocol by asking the user to solve multiple independent instances of a problem but allowing the user to make mistakes on certain fraction of problems so that the legitimate party, even though imperfect, gets accepted. The problem of proving security for such a *parallel repetition with threshold* protocol translates to our "Chernoff-Type" Direct Product Theorem.

The generic nature of weakly-verifiable puzzles allows us to extend the result easily to various other settings where we want to answer similar gap amplification questions. One such example would be to obtain a Chernoff-Type Direct Product Theorem for multi-valued functions. Another interesting setting is a secret agreement protocol with a passive eavesdropper. If the communicating parties have a higher chance (over their random tapes) of agreeing on a secret message than an eavesdropper figuring out the secret message by looking at the communication, then the theorem says that by having multiple independent communication rounds the valid parties can share more secret messages than the eavesdropper could guess. An interesting open question then is whether we can use this property to construct a secret agreement protocol where the valid parties have a high chance of agreeing on a message while the eavesdropper remains clueless about the secret message.

1.3 Security Amplification for Interactive Cryptographic Primitives

Direct Product Theorems have been used to show security amplification in cryptographic primitives such as one-way functions, collision-resistant hash functions, encryption schemes and weakly verifiable puzzles. However, there are instances [BIN97, PW07] where Direct Product Theorems do not show security amplification, specifically in protocols where the security analysis involves multiple rounds of interaction between two parties. This essentially means that Direct Product Theorems cannot be used to show security amplification for protocols in general and we need to give specific arguments to show security amplification. In a joint work [DIJK08] with Yevgeniy Dodis, Russell Impagliazzo, and Valentine Kabanets, we study security amplication of interactive cryptographic primitives, such as message authentication codes (MACs), digital signatures (SIGs) and pseudorandom functions (PRFs). This work leads us to some interesting future challenges. For instance, for the case of MACs, we show security amplification but on the cost of increasing the size of the MAC and the keys which is undesirable from a practical point of view. An interesting future direction is to show security amplification without increasing the size of the MAC and/or the keys.

1.4 A Cryptographic Channel Model

In a joint work [IJKKK07] with Russell Impagliazzo, Valentine Kabanets, Bruce M. Kapron and Valerie King we propose a very general model of channel with states, which makes fewer assumptions about the way the channel is constructed or the computational resources of the users and attackers. Much of the previous work on using communication primitives (e.g. a channel) to achieve security goals uses a *functional* model of the primitives[3]. The problem with this model is that the actual implementation gives out side information which has not been accounted for, and which an adversary can make use of. In contrast, this proposal models channels *operationally*, by a list of security and reliability properties that the channel is assumed to have, that is, by what the users and attackers can do. If the protocol is proven secure, then any attack on the protocol will yield an attack on the underlying channel that violates one of these properties. Secondly, in most information theoretic results, the functionally defined channels are used in protocols to achieve security goals which are defined operationally. This means that results in information-theoretic security do not necessarily compose. Unlike these results, reductions for the channel model proposed here securely compose, because both assumptions and conclusions are operational. We hope that many of the powerful results of information-theoretic security can be reproved in the proposed model.

The model poses some interesting hardness amplification questions. For example, the first question we are interested in is the question of *security amplification.* Here when Alice sends a random bit, Bob has a certain minimum probability of guessing the bit (reliability property) while Eve can only guess with some bounded probability

[3] For example, a broadcast channel would be modeled by the function that, on input *Send m* from any user, outputs m to all users

(security property). Given that there is a certain minimum gap between these probabilities can we amplify this gap (by some kind of repetition protocol) so that Bob almost certainly guesses the bit while Eve cannot do much better than random guessing?

2 Bounded Independence Fools Halfspaces

Halfspaces, or threshold functions, are a central class of Boolean functions $h : \{-1, +1\}^n \rightarrow \{-1, +1\}$ of the form:

$$h(x) = \mathrm{sign}(w_1 x_1 + \cdots + w_n x_n - \theta),$$

where the weights $w_1, \ldots, w_n$ and the threshold θ are arbitrary real numbers. These functions have been studied extensively in a variety of contexts like Computational Complexity, Learning Theory, and Social Choice Theory. In this work we make progress on a natural complexity-theoretic question about halfspaces. In a joint work [DGJSV09] with Ilias Diakonikolas, Parikshit Gopalan, Rocco Servedio, and Emanuele Viola, we construct the first explicit pseudorandom generators $G : \{-1, +1\}^s \rightarrow \{-1, +1\}^n$ with short seed length s that "fool" any halfspace $h : \{-1, +1\}^n \rightarrow \{-1, +1\}$, i.e. satisfy $|\mathbf{Exp}_{x \in \{-1,+1\}^s}[h(G(x))] - \mathbf{Exp}_{x \in \{-1,+1\}^n}[h(x)]| \leq \epsilon$, for a small ϵ. More specifically, we show that any k-wise independent distribution[4] fools any halfspace provided $k \geq \frac{C}{\epsilon^2} \log^2(\frac{1}{\epsilon})$, where C is some absolute constant. We then use standard explicit constructions of k-wise independent distributions over $\{-1, +1\}^n$ [CG89, ABI86] that have seed length $O(k \cdot \log n)$ to obtain explicit pseudorandom generators that fool any halfspace with error ϵ and have seed length $s = O(\log n \cdot \epsilon^{-2} \log^2(\epsilon^{-1}))$. One of the future goals is to understand the degree of independence that is required to ϵ-fool degree $d > 1$ polynomial threshold functions over $\{-1, +1\}^n$

3 Streaming Algorithms for Clustering

In joint work [AJM09] with Nir Ailon and Claire Monteleoni, we investigate Streaming and Online algorithms for Clustering problems that arise in Unsupervised Learning. For instance, the k-means and the k-medoid problem. In some recent development, we extend the k-means++ algorithm [AV07] (a very useful approximation algorithm for the k-means problem in the batch setting) to obtain a bi-criterion approximation algorithm with a better approximation guarantee with respect to the cost of the solution. We hope to use this improved algorithm to obtain better streaming algorithms for the k-means problem. This work is currently in progress.

4 Some other Projects

Here are some other projects I have been interested in.

(i) Maximum Independent Set in Degree Bounded Graphs Given a graph with bounded degree 3, we consider the problem of finding a good exact algorithm for finding a Maximum Independent Set in the graph. In a joint work [IJa] with Russell Impagliazzo, we give a $2^{n/6}$-time backtracking algorithm which was the best running time for the problem at the time of discovery[5]. In the future I hope to examine the possibility of improving our algorithm and extending our techniques to graphs with higher degree bound.

(ii) Convergence Properties of Hierarchical Markov Chains We study convergence properties of some special classes of markov chains which occur frequently in local search and randomized approximate counting. In a joint work [IJb] with Russell Impagliazzo, we give general sufficient conditions for uniform rapid convergence of the Metropolis algorithm at all temperatures on hierarchical search graphs ,where edges are between solutions of same or consecutive integer values. Such problems include graph matching and independent set. Our conditions generalize

[4] A distribution $\mathcal{D}$ on $\{-1, +1\}^n$ is called k-wise independent if the projection of $\mathcal{D}$ on any k indices is uniformly distributed over $\{-1, +1\}^k$.

[5] more recent work of Furer [F06] gives an algorithm which achieves a better bound.

the results of Jerrum and Sinclair showing rapid convergence for Metropolis on the chain of matchings of dense graphs. In the future, I hope to revisit this study and explore some other useful applications of our generalization.

References

[ABHL03] L. von Ahn, M. Blum, N. J. Hopper, J. Langford.: CAPTCHA: Using hard AI problems for security. In Advances of Cryptology (EUROCRYPT'03), pages 294–311, 2003.

[ABI86] N. Alon, L. Babai, A. Itai.: A fast and simple randomized algorithm for the maximal independent set problem. Journal of Algorithms, 7:567–583, 1986.

[AJM09] N. Ailon, R. Jaiswal, C. Monteleoni.: Streaming Algorithms for Clustering. Work in progress, 2009.

[AV07] David Arthur and Sergei Vassilvitskii.: k-means++: the advantages of careful seeding. In Proceedings of the 18th annual ACM-SIAM symposium on Discrete algorithms (SODA'07), pages 1027–1035, 2007.

[BIN97] M. Bellare, R. Impagliazzo, M. Naor.: Does parallel repetition lower the error in computationally sound protocols? In Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS'97), pages 374–383, 1997.

[CG89] B. Chor, O. Goldreich.: On the power of two point based sampling. Journal of Complexity, 5(1):96 – 106, 1989.

[CHS05] R. Canetti, S. Halevi, M. Steiner: Hardness amplification of weakly verifiable puzzles. In Proceeding of Theory of Cryptography Conference (TCC'05), pages 17–33, 2005.

[DGJSV09] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. Servedio, E. Viola.: Bounded Independence Fools Halfspaces. Submitted, 2009.

[DIJK08] Y. Dodis, R. Impagliazzo, R. Jaiswal, V. Kabanets.: Security Amplification for Interactive Cryptographic Primitives. In Proceeding of Theory of Cryptography Conference (TCC'09), pages 128–145, 2009.

[DVW09] Y. Dodis, S. Vadhan, D. Wichs.: Proofs of Retrievability via Hardness Amplification. In Proceeding of Theory of Cryptography Conference (TCC'09), pages 109–127, 2009.

[F06] M. Furer: A Faster Algorithm for Finding Maximum Independent Sets in Sparse Graphs In Proceedings of the 7th Latin American Symposium on Theoretical Informatics (LATIN'06), pages 491–501, 2006.

[I02] R. Impagliazzo. Hardness as randomness: a survey of universal derandomization. Proceedings of the ICM, Beijing 2002, 3:659672, 2002.

[IJa] R. Impagliazzo and R. Jaiswal.: A $2^{n/6}$-time algorithm for maximum independent set in graphs with bounded degree 3. Manuscript. 2005.

[IJb] R. Impagliazzo and R. Jaiswal.: General conditions for rapid convergence of metropolis on hierarchical markov chains. Manuscript. 2005

[IJK06] R. Impagliazzo, R. Jaiswal and V. Kabanets.: Approximately list-decoding direct product codes and uniform hardness amplification. In Proceeding of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06), pages 187–196, 2006. (To appear in SIAM Journal on Computing.)

[IJK07] R. Impagliazzo, R. Jaiswal and V. Kabanets.: Chernoff-type Direct Product Theorems. Journal of Cryptology 22: 75–92, 2009. Preliminary version in CRYPTO 2007, pages 500–516.

[IJKKK07] R. Impagliazzo, R. Jaiswal, V. Kabanets, B. M. Kapron, V. King.: Security Amplification for Channels with Memory. Manuscript, 2007.

[IJKW08] R. Impagliazzo, R. Jaiswal, V. Kabanets and A. Wigderson.: Uniform Direct Product Theorems: Simplified, Optimized, and Derandomized. In Proceedings of the 40th Symposium on Theory of Computing (STOC'08), pages 579–588, 2008.

[IW97] R. Impagliazzo and A. Wigderson.: $P = BPP$ if E requires exponential circuits: Derandomizing the XOR lemma. In Proceedings of the 29th Symposium on Theory of Computing (STOC'97) pages 220–229, 1997.

[JK07] A. Juels, B.S. Kaliski.: Pors: proofs of retrievability for large files. In Proceedings of the 14th ACM conference on Computer and communications security (CCS'07), pages 584–597, 2007.

[O'D04] R. O'Donnell. Hardness amplification within NP. *Journal of Computer and System Sciences*, 69(1):68–94, 2004. (preliminary version in STOC'02).

[NW94] N. Nisan and A. Wigderson.: Hardness Vs. Randomness. Journal of Computer and System Sciences, 49:149–167, 1994.

[PW07] K. Pietrzak, D. Wikstrom.: Parallel repetition of computationally sound protocols revisited. In Proceedings of Theory of Cryptography Conference (TCC'07), pages 86–102., 2007

[T03] L. Trevisan.: List-decoding using the XOR lemma. In Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS'03), pages 126–135, 2003.

[Tre05] L. Trevisan. On uniform amplification of hardness in NP. In Proceedings of the 37th Symposium on Theory of Computing (STOC'05), pages 31–38, 2005.

[Yao82] A.C. Yao. Theory and applications of trapdoor functions. In Proceedings of the 23th Annual IEEE Symposium on Foundations of Computer Science (FOCS'82), pages 80–91, 1982.