Statement of Research

KE YANG

Computer Science Department
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
(412)268-3069
yangke@cs.cmu.edu
http://www.cs.cmu.edu/~yangke

I am a Ph.D. student in Computer Science at Carnegie Mellon University. I expect to graduate on May of 2004. My research interests mainly lie in theoretical computer science, including cryptography and computer security, quantum information theory and quantum computation, machine learning, and computational complexity. Some of my research work are naturally related to several of these fields.

In below, I discuss my work in more details.

1 Cryptography and Computer Security

My work on cryptography mainly focuses on constructing protocols and cryptographic primitives such as zero-knowledge protocols, commitment schemes, oblivious transfer protocols, that enjoy a very high level of security (e.g. non-malleability, concurrent security, and/or universal composability) and are very efficient (so as to be practical) at the same time. Another part of my research is on foundations of cryptography. This includes studying the fundamental difference between programs and black boxes, with applications to the (im)possibility of obfuscating programs, and the necessity of the notion of non-malleability.

Impossibility of obfuscating programs. Intuitively, an "obfuscator" is a compiler that automatically converts any programs into an equivalent one that is "unreadable." There have been many ad hoc practices on constructing obfuscators. There is even an International Obfuscated C Code Contest. Obfuscators, if possible, would be very useful in cryptography and security, with many important applications. We investigate the problem of program obfuscation from a theoretical perspesctive. First, we formulate rigorous definitions on obfuscators that capture the intuition that obfuscators convert programs into "virtual black-boxes." We then prove that even with a very weak security formulation, program obfuscation is impossible. We do so by showing that exist functions that yield more information when written as programs than when given as an oracle. Ruling out obfuscators, our result also shows that programs, regardless how they are written, may invariantly carry strictly more information than black boxes.

This is joint work with Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, and Salil Vadhan.

Paper "**On the (Im)possibility of Obfuscating Programs**" appeared in the *21st Annual International Cryptology Conference* (CRYPTO 2001), pp. 1 – 18, LNCS 2139, 2001.

Efficient fair secure multi-party computation. Fairness in multi-party computation (MPC) protocols refers to the property that either all participating parties receive their outputs or no party receives any information at all. Obviously, fairness is very desirable for secure MPC protocols. However, it was known that under standard definitions, there does not exist fair MPC protocols when a majority of the parties are corrupted. We propose a modified security definition that allows the protocol to depend on the running time of the adversaries. The new definition fits into the standard simulation paradigm and at the same time avoids the impossibility result. We then define a "commit-prove-fair-open" functionality $\mathcal{F}_{\mathrm{CPFO}}$ and construct an efficient protocol that securely realized it, using a cryptographic primitive known as "time-lines." Finally we show that many existing (unfair) secure MPC protocols can be converted into fair protocols using the $\mathcal{F}_{\mathrm{CPFO}}$ functionality, while maintaining their security. the conversion is very efficient and preserves the security of the original protocols. As an example, we show how to use our technique to construct a very efficient protocol to solve the socialist millionaires' problem, which remains secure when arbitrarily composed with any protocols.

This is joint work with Juan Garay and Philip MacKenzie.

Paper "**Efficient and Secure Multi-Party Computation with Faulty Majority and Complete Fairness**" submitted.

Strengthening zero knowledge protocol using signatures. Zero-knowledge (ZK) protocols play an extremely important role in the field of cryptography, both theoretically and in practice. It is important to design ZK protocols that satisfy various (strong) security conditions (e.g. concurrently security, non-malleability, and/or universal composability), and remain efficient so as to be practical. We show a novel technique of using signature schemes to convert a type of weak zero-knowledge protocols (known as Σ-protocols) into concurrently secure, non-malleable, and/or universally composable. Our conversion is simple and very efficient, preserving the round of the original protocol, and has an additive overhead of constant number of exponentiations in terms of computational complexity.

This is joint work with Juan Garay and Philip MacKenzie.

Paper "**Strengthening Zero-Knowledge Protocols using Signatures**" appeared in *Eurocrypt 2003*, Warsaw, Poland, LNCS 2656, pp.177-194, 2003.

Simulation-sound trapdoor commitments. First introduced by Garay *et al.*, simulation-sound trapdoor commitments (SSTCs) are commitment schemes with the additional property that with the "trapdoor information," one can fake a commitment and open it to arbitrarily values, but without the trapdoor information, an adversary cannot open a commitment to different values, even if it sees the double opening of other commitments. SSTCs are very strong primitives and can be used to construct, for example, non-malleable and/or universally composable ZK protocols. On the other hand, SSTCs admit very efficient constructions, which allows us to construct, for example, the most efficient universally composable ZK protocols known. We present a variant of SSTC definition that are even simpler and more efficient than the previous one, yet is still powerful enough for its applications. We also investigate the relationship between SSTC schemes and non-malleable commitment schemes and prove a sequence of implication/separation results. These results in particular imply that SSTC schemes are non-malleable, which subsumes some of the previous constructions of non-malleable commitment schemes.

This is joint work with Philip MacKenzie.

Paper "**On Simulation-Sound Trapdoor Commitments** " submitted.

Efficient universally composable committed oblivious transfer. Committed oblivious transfer (COT) is a primitive that combines the bit commitment and oblivious transfer. An extended version, known as extended committed oblivious transfer (ECOT), additionally allows a party to prove relations among the committed bits. We show an efficient protocol that securely realizes the ECOT functionality and is universally composable secure. Using the ECOT as a building block, we can construct very efficient and universally composable protocols for general two-party and multi-party computation. Putting the results together, we obtain universally composable two-party and multi-party computation protocols that are much more efficient than previous ones. In particular, it only involves constant number of exponentiation operations and round per gate evaluation.

This is joint work with Juan Garay and Philip MacKenzie.

Paper "**Efficient and Universally Composable Committed Oblivious Transfer and Applications** to appear in *Theory of Cryptography Conference (TCC '04)*, Cambridge, MA, 2004.

Alternative and more efficient definitions of non-malleability. Non-malleable encryption is a very important notion in modern cryptography. Motivated by real-world examples, non-malleability intuitively means that observing an encrypted message does not help an adversary produce an encryption of a "related" message. Known constructions for non-malleable encryptions tend to be more complex than the ones for semantically secure encryptions. We investigate the question whether non-malleability is indeed necessary and furthermore, whether there exist alternative definitions to non-malleability that are sufficient for their application and admit simpler constructions. We answer the question positively by presenting two alternative definitions. Both of the alternative definitions have simpler constructions, and both find applications in situations that are previously believed to need non-malleability.

This is joint work with Philip MacKenzie and Mike Reiter.

Paper "**Alternatives to Non-Malleability: Definitions, Constructions and Applications** to appear in *Theory of Cryptography Conference (TCC '04)*, Cambridge, MA, 2004.

My work on computer security include private information retrieval, anonymous communication and formal verification of authentication protocols.

Private information retrieval with applications in anonymous communication. We build a variant of the private information retrieval (PIR) system that additionally allows keyword search with access control, using a threshold homomorphic cryptosystem as the underline infrastructure. One salient property of our system is that it is very efficient, and in particular, the communication complexity is independent from the number of records stored. As an interesting example, we use this system to build an efficient and unlinkable anonymous messaging service, which can be easily modified into both sender and receiver-anonymous.

This is joint work with Lea Kissner, Alina Oprea, Mike Reiter, and Dawn Song.

Paper "**Private Push and Pull with Applications to Anonymous Communication**" submitted.

Formal verification of authentication protocols. We build a system that automatically verifies the correctness of authentication protocols. When an insecurity is discovered, the system will generate an attack scenario that breaks the protocol. Our system is written in Java and is very efficient.

This is joint work with Liansheng Huang, Xinbing Wang, Feng Xie.

Paper "**Formal Authentication Based on Intruder of Role Impersonate**" (in Chinese) appeared in *Journal of Tsinghua University*, July, 2001.

2 Quantum Information Theory and Quantum Computation

My work on quantum information theory focuses on entanglement distillation protocols (EDPs).

Entanglement is arguably the quintissential feature that distinguishes quantum information theory from classical information theory. In its purest form, entanglement appears as EPR pairs. Entanglement and EPR pairs play an important role in almost all aspects in quantum computation and quantum information theory, including teleportation, Shor's algorithm for factoring and the unbreakable quantum code. However, researchers still don't have a complete understand on entanglement and EPR pairs. In particular, it was not clear how to quantify the "amount" of entanglement of a system. It wasn't clear either, that if "entanglement" is a physical quantity that is preserved in operation, like energy and momentum.

EDPs are two-party protocols that "distills" pure entanglement from imperfect entanglemnt. More precisely, they take partially entangled qubits as inputs, perform local (quantum) operations and classical communications, and output near-perfect EPR pairs. They are also known as "entanglement purification protocols." EDPs are of extreme importance in quantum information theory, since EPR pairs are extremely sensitive to noise and degrade very quickly. It is essential to employ entanglement distillation protocols to "distill" pure EPR pairs in order to defeat the noise.

EDPs with limited information. We study entanglement distillation protocols in a very general setting, where the only information available is the fidelity of the input state. We present a protocol which is essential optimal in this model. We also prove lower bounds on how well a protocol can do.

This is joint work with Andris Ambainis and Adam Smith.

Paper "**Extracting Quantum Entanglement (General Entanglement Purification Protocols**" appeared in the *IEEE Conference of Computational Complexity (CCC 2002)*, Montréal, Quebéc, Canada, pp. 103-112, 2002.

Classical communication complexity of EDPs. We study the classical communication complexity of these protocols for various "noise models," which describe how the EPR pairs are corrupted.

This is joint work with Andris Ambainis.

Paper "**Towards the Classical Communication Complexity of Entanglement Distillation Protocols with Incomplete Information**" in LANL eprint `quant-ph/0207090`.

My work on quantum computation is on efficient NMR quantum computing.

Efficient NMR quantum computing. With the current technology, NMR appears to be the most promising approach to build quantum computers. However, the model for NMR computing is different from the "standard model." Therefore one needs to translate quantum algorithms in the standard model into ones in the NMR model. Known general translation technique is not very efficient in that it bring a large overhead in the complexity of the programs. We ask the question whether more efficient general translations exist, and answer it in the negative by exhibiting some impossibility results (both information theoretic and computational). Our results use techniques from machine learning and cryptography.

This is joint work with Avrim Blum.

Paper **On Statistical Query Sampling and NMR Quantum Computing** appeared in *18th Annual IEEE Conference of Computational Complexity (CCC 2003)*, Århus, Denmark, pp. 194-205, 2003.

3 Machine Learning

My work on machine learning focuses on a specific model, known as the *statistical query (SQ)* model. In the SQ model, a learning algorithm gathers information about the target concept by submitting predicates to an "SQ oracle," which then returns an estimate of the probability that the predicates being true. The SQ model is an obvious restriction to the commonly used probably almost correct (PAC) model, but nevertheless proved to be very useful, due to its inherent noise-tolerance property. In fact, it was shown that a large fraction of known PAC learning algorithm can be casted as a SQ algorithm. Naturally, it is important to understand the inherent limitation of SQ algorithms, and in particular, to find examples of concept classes that separate the SQ model from the PAC model.

We proved various lower bounds on the SQ learning that separate the SQ model from the PAC model. These results improve previous ones and some novel techniques used in the proof can be interesting by themselves.

Learning correlated functions using SQ. We show a lower bound on SQ for learning a concept class where any pair of concepts are correlated. This is a natural extension to a previous result, which is concerns with *orthogonal* concept class where the concepts are completely uncorrelated. Our lower bound is almost tight. This result, when applied on a specific concept class known as "booleanized linear functions," has implications on the security of certain cryptosystem, and yields an interesting contrast to learning perceptrons. We also gave a PAC algorithm that learns the class of booleanized linear functions more efficiently than SQ, therefore exhibiting the first separation results between SQ and PAC for concept classes that are not parity functions. The PAC algorithm utilizes a technique that might be interesting by itself.

Paper "**On Learning Correlated Boolean Functions Using Statistical Query**" appeared in the *Twelfth International Conference on Algorithmic Learning Theory (ALT 2001)*, Washington, DC, LNAI 2225, pp. 59-76, 2001.

More Lower bounds on SQ. In this work, we prove two more lower bounds on the SQ model. The first one both improves and extends my previous work on learning correlated concept classes by proving a better lower bound (that matches known upper bounds up a logarithmic factor) for a stronger variant of SQ-learning algorithms. The second result is concerned with learning an arbitrary concept class and uses an interesting technique that exploits results from singular value decomposition (SVD).

Paper "**New Lower Bounds for Statistical Query Learning**" appeared in the *Fifteenth Annual Conference on Computational Learning Theory (COLT 2002)*, Sydney, NSW, Australia, LNAI 2375, pp. 229-243, 2002.

4 Computational Complexity

My work on computational complexity various topics such as circuit complexity and communication complexity.

Impossibility of non-interactive correlation distillation. Suppose Alice and Bob each has as private input a collections of bits (denoted as $\{a_i\}$ and $\{b_i\}$, respectively) that are somewhat related. More precisely, we have $\Pr[a_i = b_i] = \lambda$ for some $\lambda > 1/2$. Now they wish to each apply a mapping to their inputs outputting one bit, denoted as X and Y. They wish to maximize $\Pr[X = Y]$ while insisting that both X and Y are unbiased. How well can they do?

This problem is known as non-interactive correlation distillation and is closely related to topics in computer science including information reconciliation and random beacons. We prove two negative results that Alice and Bob cannot increase the correlation of their outputs arbitrarily close to one. We also construct a one-bit protocol that allows Alice and Bob to achieve higher correlation than what is possible without interaction. The protocol shows that even the minimal amount of communication is provably more powerful than no communication at all.

Paper "**On the (Im)possibility of Non-interactive Correlation Distillation**" to appear in *Latin American Theoretical INformatics (LATIN 2004)*, Buenos Aires, Argentina, 2004.

Computational complexity of integer circuit evaluation. Integer circuits are combinatorial circuits where on the wires are sets of natural numbers, and the gates are union, pair-wise addition and pair-wise multiplication. The integer circuit evaluation (ICE) problem is given a circuit with its inputs, to determine if a certain integer x is contained in the output wire (which is an integer set). Integer circuits are a natural extension to the arithmetic circuit and have been studied by various researchers. It was known that ICE is in PSPACE but not known if it is PSPACE-complete. We proved that ICE is indeed PSAPCE-complete by showing a reduction from the quantified boolean formula (QBF) to ICE.

Paper "**Integer Circuit Evaluation is PSPACE-complete**" appeared in the *IEEE Conference of Computational Complexity (CCC 2000)*, Florence, Italy, pp. 204 - 213, 2000.

Computational complexity of MAX/MIN/AVRG circuits. MAX/MIN/AVRG circuits have real values between 0 and 1 on the wires and MAX, MIN, and AVERAGE gates of fan-in 2. Furthermore, these circuits may have feed-backs. MAX/MIN/AVRG circuits are interesting because they can be used to model two-person stochastic games and Markov decisional process. It can be shown that each circuit has a "stable" solution whereas all the gates are satisfied. Nevertheless, it is not know if one can find a stable solution efficiently. We study the complexity of the MAX/MIN/AVRG circuits and discuss various results, as well as applications.

This is joint work with Manuel Blum and Rachel Rue.

Paper "**On the Complexity of MAX/MIN/AVRG Circuits**" published as *CMU SCS Technical Report, CMU-CS-02-110*, 2002.

Complexity of data expansion for secret sharing using XOR. Secret sharing was first introduced by Shamir and soon became an extremely important tool in various areas in computer science. A (t, n) sharing scheme splits a secret into n shares, such that from any t shares one can reconstruct the secret, but any $(t-1)$ shares yield no information about the secret at all. General (t, n) sharing schemes with constant data expansion factor exist, but they normally involve field operations and can inefficient. On the other hand, there exists very efficient $(2, n)$ sharing schemes where the operations only involves XOR. But these XOR-based schemes have larger data expansion factors. We prove a lower bound on the data expansion factor for XOR based $(2, n)$ sharing scheme. We also construct a scheme whose expansion factor matches this lower bound.

This is joint work with Xinliang Lin and Yiqi Dai.

Paper "**Minimal Size of $(2, n)$) Data Sharing Scheme Under XOR Operation** (in Chinese) appeared in *Journal of Tsinghua University*, May, 1998.

List of Papers (reverse chronological)

1. J. Garay, P. MacKenzie, *K. Yang*.
   Efficient and Secure Multi-Party Computation with Faulty Majority and Complete Fairness.
   Submitted.

2. L. Kissner, A. Oprea, M. Reiter, D. Song, *K. Yang*.
   Private Push and Pull with Applications to Anonymous Communication .
   Submitted.

3. P. MacKenzie, *K. Yang*.
   On Simulation-Sound Trapdoor Commitments.
   Submitted.

4. *K. Yang*.
   On the (Im)possibility of Non-interactive Correlation Distillation.
   To appear in *Latin American Theoretical INformatics (LATIN 2004)*, Buenos Aires, Argentina, 2004.

5. J. Garay, P. MacKenzie, *K. Yang*.
   Efficient and Universally Composable Committed Oblivious Transfer and Applications.
   To appear in *Theory of Cryptography Conference (TCC '04)*, Cambridge, MA, 2004.

6. P. MacKenzie, M. Reiter, *K. Yang*.
   Alternatives to Non-Malleability: Definitions, Constructions and Applications.
   To appear in Theory of Cryptography Conference (TCC '04), Cambridge, MA, 2004.

7. A. Blum, *K. Yang*.
   On Statistical Query Sampling and NMR Quantum Computing.
   Appeared in *18th Annual IEEE Conference of Computational Complexity (CCC 2003)*, Århus, Denmark, pp. 194-205, 2003.

8. J. Garay, P. MacKenzie, *K. Yang*.
   Strengthening Zero-Knowledge Protocols using Signatures.
   In *Eurocrypt 2003*, Warsaw, Poland, LNCS 2656, pp.177-194, 2003.

9. A. Ambainis, *K. Yang*.
   Towards the Classical Communication Complexity of Entanglement Distillation Protocols with Incomplete Information.
   In *LANL eprint* report number `quant-ph/0207090`.

10. *K. Yang*.
    New Lower Bounds for Statistical Query Learning.
    Appeared in the *Fifteenth Annual Conference on Computational Learning Theory (COLT 2002)*, Sydney, NSW, Australia, LNAI 2375, pp. 229-243, 2002.

11. M. Blum, R. Rue, *K. Yang*.
    On the Complexity of MAX/MIN/AVRG Circuits.
    CMU SCS Technical Report, CMU-CS-02-110, 2002.

12. A. Ambainis , A. Smith, *K. Yang*.
    Extracting Quantum Entanglement (General Entanglement Purification Protocols).
    Appeared in the *IEEE Conference of Computational Complexity (CCC 2002)*, Montréal, Quebéc, Canada, pp. 103-112, 2002.

13. *K. Yang*.
    On Learning Correlated Boolean Functions Using Statistical Query.
    Appeared in the *Twelfth International Conference on Algorithmic Learning Theory (ALT 2001)*, Washington, DC, LNAI 2225, pp. 59-76, 2001.

14. B. Barak, O. Goldreich, R. Impagliazzo, S. Rudich, A. Sahai, S. Vadhan, *K. Yang*.
On the (Im)possibility of Obfuscating Programs.
In the *21st Annual International Cryptology Conference (CRYPTO 2001)*, pp. 1 – 18, 2001.

15. L. Huang, X. Wang, F. Xie, *K. Yang*.
Formal Authentication Based on Intruder's Role Impersonate (in Chinese).
Appeared *Journal of Tsinghua University*, July, 2001.

16. *K. Yang*.
Integer Circuit Evaluation is PSPACE-complete.
Appeared in the *IEEE Conference of Computational Complexity (CCC 2000)*, Florence, Italy, pp. 204 - 213, 2000. Journal version at *Journal of Computer and System Sciences*, 63, 288–303 (2001).

17. *K. Yang*, X. Lin, Y. Dai
Minimal Size of $(2, n)$ Data Sharing Scheme Under XOR Operation (in Chinese).
Appeared in *Journal of Tsinghua University*, May, 1998.

List of Abstracts (reverse chronological)

1. J. Garay, P. MacKenzie, *K. Yang*.
Efficient and Secure Multi-Party Computation with Faulty Majority and Complete Fairness.

Abstract We study the problem of constructing secure multi-party computation (MPC) protocols that are *completely fair* — meaning that either all the parties learn the output of the function, or nobody does — even when a majority of the parties are corrupted. We first propose a framework for fair multi-party computation, within which we formulate a definition of secure and fair protocols. The definition follows the standard simulation paradigm, but is modified to allow the protocol to depend on the running time of the adversary. In this way, we avoid a well-known impossibility result for fair MPC with corrupted majority; in particular, our definition admits constructions that tolerate up to $(n-1)$ corruptions, where n is the total number of parties. Next, we define a "commit-prove-fair-open" functionality and construct an efficient protocol that realizes it, using a new variant of a cryptographic primitive known as "time-lines." Finally, we show that some of the existing secure MPC protocols can be easily transformed into fair protocols by using the "commit-prove-fair-open" functionality. Putting these results together, we construct efficient, secure MPC protocols that are completely fair even in the presence of corrupted majorities. Furthermore, these protocols remain secure when arbitrarily composed with any protocols, which means, in particular, that they are concurrently-composable and non-malleable. Finally, as an example, we show a very efficient protocol that fairly and securely solves the socialist millionaires' problem.

2. L. Kissner, A. Oprea, M. Reiter, D. Song, *K. Yang*.
Private Push and Pull with Applications to Anonymous Communication.

Abstract We propose a modification of the Private Information Retrieval (PIR) model to allow modification of the database from which information is requested and rich searching on keywords, with access control. To accomplish this, we give Private Push and Pull (P^3) protocols for a group of n servers. The communication complexity between the client and the servers is independent of the number of records in the database (or more generally, the number of previous push and pull transactions) and can be independent on the number of servers, depending on the choice of cryptographic primitives. Our scheme relies on a partially homomorphic cryptosystem, for which there is an algorithm for composing ciphertexts to get an encryption of the added plaintexts. To demonstrate the utility of P^3 , we use it to implement an unlinkable anonymous communication message service, which can easily be extended to one that provides both sender and receiver anonymity.

3. P. MacKenzie, *K. Yang*.
On Simulation-Sound Trapdoor Commitments.

Abstract We study the recently introduced notion of a *simulation-sound trapdoor commitment (SSTC)* scheme. In this paper, we present a new, simpler definition for an SSTC scheme that admits more efficient constructions and can be used in a larger set of applications. Specifically, we show how to construct SSTC schemes from any one-way functions, and how to construct very efficient SSTC schemes based on specific number-theoretic assumptions. We also show how to construct simulation-sound, non-malleable, and universally-composable zero-knowledge protocols using SSTC schemes, yielding, for instance, the most efficient universally-composable zero-knowledge protocols known. Finally, we explore the relation between SSTC schemes and non-malleable commitment schemes by presenting a sequence of implication and separation results, which in particular imply that SSTC schemes are non-malleable.

4. *K. Yang*.
On the (Im)possibility of Non-interactive Correlation Distillation.

Abstract We study the problem of non-interactive correlation distillation (NICD). Suppose Alice and Bob each has a string, denoted by $A = a_0 a_1 \cdots a_{n-1}$ and $B = b_0 b_1 \cdots b_{n-1}$, respectively. Furthermore, for every $k = 0, 1, ..., n-1$, (a_k, b_k) is independently drawn from a distribution $\mathcal{N}$, known as the "noise mode". Alice and Bob wish to "distill" the correlation non-interactively, i.e., they wish to each apply a function to their strings, and output one bit, denoted by X and Y, such that $\Pr[X = Y]$ can be made as close to 1 as possible. The problem is, for what noise model can they succeed? This problem is related to various topics in computer

science, including information reconciliation and random beacons. In fact, if NICD is indeed possible for some general class of noise models, then some of these topics would, in some sense, become straightforward corollaries.

We prove two negative results on NICD for various noise models. We prove that for these models, it is impossible to distill the correlation to be arbitrarily close to 1. We also give an example where Alice and Bob can increase their correlation with one bit of communication. This example, which may be of its own interest, demonstrates that even the smallest amount of communication is provably more powerful than no communication.

5. J. Garay, P. MacKenzie, *K. Yang*.
Efficient and Universally Composable Committed Oblivious Transfer and Applications.

Abstract Committed Oblivious Transfer (COT) is a useful cryptographic primitive that combines the functionalities of bit commitment and oblivious transfer. In this paper, we introduce an extended version of COT (ECOT) which additionally allows proofs of relations among committed bits, and we construct an efficient protocol that securely realizes an ECOT functionality in the universal-composability framework. Then, using the ECOT functionality as a building block, we construct universally-composable protocols for general two-party and multi-party functionalities. Our constructions are more efficient than previous ones, involving only a constant number of exponentiations and a constant number of communication rounds per gate evaluation.

6. P. MacKenzie, M. Reiter, *K. Yang*.
Alternatives to Non-Malleability: Definitions, Constructions and Applications.

Abstract We explore whether non-malleability is necessary for the applications typically used to motivate it, and propose two alternatives. The first we call weak non-malleability (wnm) and show that it suffices to achieve secure contract bidding (the application for which non-malleability was initially introduced), despite being strictly weaker than non-malleability. The second we call tag-based non-malleability (tnm), and show that it suffices to construct an efficient universally-composable secure message transmission (SMT) protocol, for which the only previous solution was based on a public key encryption functionality whose security is equivalent to non-malleability. We also demonstrate constructions for wnm and tnm encryption schemes that are simpler than known constructions of non-malleable encryption schemes.

7. A. Blum, *K. Yang*.
On Statistical Query Sampling and NMR Quantum Computing.

Abstract We introduce a "Statistical Query Sampling" model, in which the goal of an algorithm is to produce an element in a hidden set $S \subseteq \{0,1\}^n$ with reasonable probability. The algorithm gains information about S through oracle calls (statistical queries), where the algorithm submits a query function $g(\cdot)$ and receives an approximation to $\Pr_{x \in S}[g(x) = 1]$. We show how this model is related to NMR quantum computing, in which only statistical properties of an ensemble of quantum systems can be measured, and in particular to the question of whether one can translate standard quantum algorithms to the NMR setting without putting all of their classical post-processing into the quantum system. Using Fourier analysis techniques developed in the related context of *statistical query learning*, and techniques in cryptography, we prove a number of lower bounds (both information-theoretic and cryptographic) on the ability of algorithms to produces an $x \in S$, even when the set S is fairly simple. These lower bounds point out a difficulty in efficiently applying NMR quantum computing to algorithms such as Shor's and Simon's algorithm that involve significant classical post-processing. We also explicitly relate the notion of statistical query sampling to that of statistical query learning.

8. J. Garay, P. MacKenzie, *K. Yang*.
Strengthening Zero-Knowledge Protocols using Signatures.

Abstract Recently there has been an interest in zero-knowledge protocols with stronger properties, such as concurrency, unbounded simulation soundness, non-malleability, and universal composability. In this paper we show a new technique that uses a signature scheme that is existentially unforgeable against adaptive chosen message attacks to construct zero-knowledge protocols with these stronger properties in the common reference string model. For instance, using our technique we transform any Σ-protocol (which is honest-verifier zero-knowledge) into an unbounded simulation sound concurrent zero-knowledge protocol. We also introduce a

variant of Σ-protocols for which our technique further achieves the properties of non-malleability and/or universal composability. In addition to its conceptual simplicity, a main advantage of this new technique over previous ones is that it allows for very efficient instantiation based on the security of some efficient signature schemes and standard number-theoretic assumptions. For instance, one instantiation of our technique yields an unbounded simulation sound zero-knowledge protocol under the Strong RSA assumption, incurring an overhead of a small constant number of exponentiations, plus the generation of two signatures.

9. A. Ambainis, *K. Yang*.
Towards the Classical Communication Complexity of Entanglement Distillation Protocols with Incomplete Information.

Abstract Entanglement is an essential resource for quantum communication and quantum computation, similar to shared random bits in the classical world. Entanglement distillation extracts nearly-perfect entanglement from imperfect entangled state. The classical communication complexity of these protocols is the minimal amount of classical information that needs to be exchanged for the conversion. In this paper, we focus on the communication complexity of protocols that operate with *incomplete information*, i.e., where the inputs are mixed states and/or prepared adversarially.

We study three models of imperfect entanglement, namely, the bounded measurement model, the depolarization model, and the fidelity model. We describe there models as well as the motivations for studying them. For the bounded measurement model and the depolarization model, we prove tight and almost-tight bounds on the output quality of non-interactive protocols. For the fidelity model we prove a lower bound that matches the upper bound given by Ambainis, Smith, and Yang and thus completely characterizes communication complexity of entanglement distillation protocols for this model. Our result also implies the optimality of the BB84 protocol in terms of communication complexity.

We emphasize that although some of the results appear intuitively straightforward, their proofs are not. In fact, two novel techniques are developed for proving these results. We believe that these techniques are of independent interests, too.

10. *K. Yang*.
New Lower Bounds for Statistical Query Learning.

Abstract We prove two lower bounds on the Statistical Query (SQ) learning model. The first lower bound is on weak-learning. We prove that for a concept class of SQ-dimension d, a running time of $\Omega(d/\log d)$ is needed. The SQ-dimension of a concept class is defined to be the maximum number of concepts that are "uniformly correlated", in that each pair of them have nearly the same correlation. This lower bound matches the previously known upper bound, up to a logarithmic factor. We prove this lower bound against an "honest SQ-oracle", which gives a stronger result than the ones against the more frequently used "adversarial SQ-oracles". The second lower bound is more general. It gives a continuous trade-off between the "advantage" of an algorithm in learning the target function and the number of queries it needs to make, where the advantage of an algorithm is the probability it succeeds in predicting a label minus the probability it does not. Both lower bounds extend and/or strengthen previous results.

11. M. Blum, R. Rue, *K. Yang*.
On the Complexity of MAX/MIN/AVRG Circuits.

Abstract We study the complexity of a class of circuits, namely, the MAX/MIN/AVRG circuits. On the wires of these circuits are real values between 0 and 1; the functions each gate performs are MAX, MIN, and AVERAGE of fan-in 2; there can be feed-backs in the circuit. It can be shown that every such circuit has at least a "stable" solution, meaning that there is a way to set each wire to a particular value such that each gate is satisfied. However, finding a stable solution in polynomial time seems to be a tricky problem and remains unsolved. We discuss some results concern this computation model, as well as its applications.

12. A. Ambainis , A. Smith, *K. Yang*.
Extracting Quantum Entanglement (General Entanglement Purification Protocols).

Abstract We study the problem of extracting EPR pairs from a general source of entanglement. Suppose Alice and Bob share a bipartite state ρ which is "reasonably close" to perfect EPR pairs. The only information Alice and Bob possess is a lower bound on the fidelity of ρ and a maximally entangled state. They wish to "purify" ρ using local operations and classical communication and output a state that is arbitrarily close to EPR pairs. We prove that on average, Alice and Bob cannot increase the fidelity of the input state significantly. On the other hand, there exist protocols that may fail with a small probability, and otherwise will output states arbitrarily close to EPR pairs with very high probability. These protocols come from the "purity-testing protocols" of Barnum et al.

13. *K. Yang*.
On Learning Correlated Boolean Functions Using Statistical Query.

Abstract In this paper, we study the problem of using statistical query (SQ) to learn a class of highly correlated boolean functions, namely, a class of functions where any pair agree on significantly more than 1/2 fraction of the inputs. We give an almost-tight bound on how well one can approximate all the functions without making any query, and then we show that beyond this bound, the number of statistical queries the algorithm has to make increases with the "extra" advantage the algorithm gains in learning the functions. Here the advantage is defined to be the probability the algorithm agrees with the target function minus the probability the algorithm doesn't agree.

An interesting consequence of our results is that the class of booleanized linear functions over a finite field ($f_{\vec{a}}(\vec{x}) = 1$ iff $\phi(\vec{a} \cdot \vec{x}) = 1$, where ϕ is an arbitrary boolean function the maps any elements in GF_p to ± 1) is not efficiently learnable. This result is useful since the hardness of learning booleanized linear functions over a finite field is related to the security of certain cryptosystem. In particular, we prove that the class of linear threshold functions over a finite field ($f_{\vec{a},b}(\vec{x}) = 1$ iff $\vec{a} \cdot \vec{x} \geq b$) cannot be learned efficiently using statistical query. This contrasts with Blum et al.'s result that linear threshold functions over reals (perceptions) are learnable using the SQ model.

Finally, we describe a PAC-learning algorithm that learns a class of linear threshold functions in time that is provably impossible for statistical query algorithms. With properly chosen parameters, this class of linear threshold functions become an example of PAC-learnable, but not SQ-learnable functions that are not parity functions.

14. B. Barak, O. Goldreich, R. Impagliazzo, S. Rudich, A. Sahai, S. Vadhan, *K. Yang*.
On the (Im)possibility of Obfuscating Programs.

Abstract Informally, an *obfuscator* O is an (efficient, probabilistic) "compiler" that takes as input a program (or circuit) P and produces a new program $O(P)$ that has the same functionality as P yet is "unintelligible" in some sense. Obfuscators, if they exist, would have a wide variety of cryptographic and complexity-theoretic applications, ranging from software protection to homomorphic encryption to complexity-theoretic analogues of Rice's theorem. Most of these applications are based on an interpretation of the "unintelligibility" condition in obfuscation as meaning that $O(P)$ is a "virtual black box," in the sense that anything one can efficiently compute given $O(P)$, one could also efficiently compute given oracle access to P.

In this work, we initiate a theoretical investigation of obfuscation. Our main result is that, even under very weak formalizations of the above intuition, obfuscation is impossible. We prove this by constructing a family of functions F that are *inherently unobfuscatable* in the following sense: there is a property $\pi : F \to \{0,1\}$ such that (a) given *any program* that computes a function $f \in F$, the value $\pi(f)$ can be efficiently computed, yet (b) given *oracle access* to a (randomly selected) function $f \in F$, no efficient algorithm can compute $\pi(f)$ much better than random guessing.

We extend our impossibility result in a number of ways, including even obfuscators that (a) are not necessarily computable in polynomial time, (b) only *approximately* preserve the functionality, and (c) only need to work for very restricted models of computation (TC_0). We also rule out several potential applications of obfuscators, by constructing "unobfuscatable" signature schemes, encryption schemes, and pseudorandom function families.

15. L. Huang, X. Wang, F. Xie, *K. Yang*.
Formal Authentication Based on Intruder's Role Impersonate (in Chinese).

Abstract The popularity of computer networks has made network security a critical issue of nowadays life. This paper introduces some formal authentication tools for protocol security, and implements an authentication protocol verification algorithm based on intruder's role impersonate. The algorithm is implemented using Java. The system is based on the idea that an authentication protocol must have the parties exchange secrets of some form, and we formulate a framework, on which we implement our algorithms.

16. *K. Yang*.
Integer Circuit Evaluation is PSPACE-complete.

Abstract In this paper, we address the problem of evaluating the Integer Circuit (IC), or the $\{\cup, \times, +\}$-circuit over the set of natural numbers. The problem is a natural extension to the integer expression by Stockmeyer and Mayer, and is also studied by Mckenzie, Vollmer and Wagner in their "Polynomial Replacement System". We show a polynomial-time algorithm that reduces QBF (Quantified Boolean Formula) problem to the Integer Circuit problem. This complements the result of Wagner to show that IC problem is PSPACE-complete. The proof in this paper provides a new perspective to describe PSPACE-completeness.