SaTC: Medium: Distribution-Matching Encryption

Ari Juels, Thomas Ristenpart, Thomas Shrimpton

Solicitation: `http://www.nsf.gov/pubs/2014/nsf14599/nsf14599.htm`

1 Introduction

Contemporary encryption schemes are almost exclusively *distribution-agnostic*. Their security properties are independent of the statistical characteristics of plaintexts and they always yield ciphertexts that are uniformly distributed in the view of an adversary. Distribution-agnostic cryptography is conceptually simple and often convenient in practice. There is a growing, unaddressed need, however, for rigorously studied cryptographic primitives that are specifically tailored to non-uniform plaintext and/or ciphertext distributions, a notion that we call *distribution-matching encryption* (DME). This need arises in practice in applications ranging from censorship resistance to password-based encryption.

We propose to unify disparate strands of research into DME within a single formal framework. This framework will yield new insights, permit a cross-pollination of concepts and techniques, and spawn new tools and techniques for a range of practical DME applications. By combining formal definitions and security proofs with empirical study, and by embellishing systems we have already built and/or fielded, we will publicly deploy the results of our research.

The Opportunity A number of applications call for DME schemes dictated by a particular operational environment. Researchers have begin to explore foundational principles for DME, but until recently, their focus has not been on application-specific settings. Steganography, an instance of DME, illustrates the gap between foundations and practice: Techniques with provable security guarantees fall short of practical efficiency [**?**], while fast, practical schemes, e.g., [**?**,**?**] lack rigorous security guarantees. Some target DME applications on which we will focus in this proposal include:

- *Censorship resistance:* Bypassing the deep-packet-inspection (DPI) traffic filtering conducted by repressive regimes requires ciphertexts that are statistically similar to permissible traffic, i.e., steganography;
- *Password-based encryption*: Several recent proposals [] have shown that the weak passwords typically chosen by users can be partly compensated for using decoy plaintexts that look statistically like valid ones;
- *Hardening human-generated secrets*: Typographical and reading errors introduce noise into passwords and biometrics. Cryptographic constructions that are sensitive to the statistical distributions over such secrets, such as secure sketches [] (entropy-bounding error-correcting codes), can mitigate side effects of noise such as decryption failures.

By exploring these applications individually we will aim to design practical new primitives with formal foundations. Through combined exploration, we will pursue a unified perspective that guides new techniques and applications and improves existing ones. We will inform this work with experimental exploration and evaluation in an *empirically-driven, provable security* approach to application design.

Our Proposal: Foundations We plan to unify disparate trains of research into DME-related concepts into a single, formal framework. To do so, we will build on the concept of a *distribution-transforming encoder* (DTE), the linchpin of the honey encryption construction in [].

A DTE is a pair $\mathsf{DTE} = (\mathsf{encode}, \mathsf{decode})$ of algorithms, where encode is a mapping from a message space $\mathcal{M}$ to a space $\mathcal{S}$ of what are called *seeds*, and decode maps seeds to messages. In the original DTE definition, a good DTE is one in which, when seeds are selected uniformly at random and encoded, they yield a distribution close to a target one p_m over $\mathcal{M}$. By extending this definition to non-uniform seed spaces and expanding the accompanying security definitions, it is possible to encompass honey encryption, FTE, fuzzy extractors, steganography, and other primitives within the same definitional framework.

A range of new steganographic constructions then come to light that we will explore in this proposal. First, by enabling information-theoretic security in applications that use low-entropy secrets (such as passwords), honey encryption points the way toward: (1) Password-based steganography and (2) Steganography for password-based key-agreement protocols. Additionally, as we explain below, a special class of DTEs offers a path in some settings to: (3) Highly efficient yet provably secure steganography.

This expanded view of DTEs also brings into view another foundational connection, between DTE constructions and techniques from simulation-based proofs of security. These proofs often involve "cooking" values returned by an oracle to create an environment that is statistically indistinguishable by an adversary from a real environment—a goal analogous to that in DME settings. In preliminary work, we have found that such techniques, when imported into DTE construction, can substantially improve efficiency in a key class of DTEs (based on rejection sampling). This approach to DTE design promises more efficient honey encryption and steganographic schemes.

Experiments and Deployment We will inform our explorations into foundations and new constructions with empirical study of deployment environments and the construction of practical tools. The starting point for this work will be statistical evaluation and modeling of packets captured in a large-scale traffic measurement study. We will aim to design efficient DTEs for empirically observed traffic that can be parlayed into techniques for circumvention of statistical DPI filtering on the internet. Building on our existing, deployed DPI-circumvention tools, we publicly release novel tools resulting from the research in this proposal.

The DTEs in honey encryption cause silent decryption failures when keys contain errors, e.g., when there are typos in passwords. Drawing on real-world password and password-vault breach data we will explore practical improvements to secure sketches that can detect typos in a client with minimal degradation in password entropy. In preliminary work, we have observed structure (high average edit distance) in popular passwords in the wild, an observation that we will leverage to construct a novel form of secure sketch. We will deploy this code in conjunction with SweetPass, a honey-encryption-based password vault under development. We will aim to extend this line of work to biometrics.

Team

Organization