Crypto Dark Matter++

Abstract

Abstract goes here.

1 Introduction

1.1 Applications

2 Background

3 Related Work

4 New Protocol

5 New Protocol - Additional Details

6 Implementation

6.1 Preliminaries

We begin by defining some basic notation that we will use throughout this work. We use uppercase letters (e.g., A, B) to denote matrices and bold letters to denote vectors $\mathbf{a}$. As was specified in the Boneh et al work [1], the input variable to the PRF functions are the key K which is of size $m \times n$ and each input $\mathbf{i}$ is a vector of size n. To save storage space, the key was implemented as a Toeplitz matrix, requiring $2 \cdot n$ bits of storage space. In phase 3 of the algorithm, a randomization matrix R is used which is of size $t \times m = 81 \times 256$, resulting in entropy of 128 bits (as $81 = 128 \times log_2 3$).

6.2 representing Z_2 vector

The protocols include a few stages, where the beginning ones handle binary data. By default, each element is therefore part of a set of 0,1. However, machine words can hold up to 64 bits. To take advantage of this property, we utilized bit slicing optimization. This bit-wise packing technique was used to optimize the run-time, each 64 bits were represented by a word. Since each key $\mathbf{K}$ is of size $m \times n$ and each input is a vector of size n, it is possible to pack each 64 rows into m words. This may result in time saving of up to $\times 64$ of the run-time.

6.3 Representing Z_3 vector

In the last stage of the PRF protocols, the binary vector is represented as a trinary vector. It is then multiplied by a matrix R in Z_3 to produce output randomized vector in Z_3 with entropy of 128 bits. To optimize the randomization in Z_3, which requires a matrix-vector multiplication in the last phase of the PRF calculation, we implement two methods:

6.3.1 Bit slicing and operations:

Bit slicing is implemented by representing a vector over Z_3 as two binary vectors - one LSB's and one MSB's.

The addition, subtraction, multiplication and MUX mod 3 were implemented using the protocols defined in table reftab:multicol. We took advantage of the fact that when representing each vector as MSB and LSB, negation (which is the same as multiplying by two) is the same as swapping the most and least significant bits. Therefore, we get:

mult-by-2 mod 3 $\leftrightarrow$ negation mod 3 $\leftrightarrow$ swap the MSB and LSB

Such a vector is sometimes being represented as a vector of $2 \times n$ bits. One example is when this vector is sent as an input to the Oblivious Transfer mechanism. The algorithms can be found in **????????**.

input: $\vec{l_1}, m_1$ - LSB and MSB vectors of value 1
l_2, m_2 - LSB and MSB vectors of value 2
s - binary select vector

Table 1: Operations in Z_3

Operations	Methods
Addition	$t := (l_1 \wedge m_2) \oplus (l_2 \wedge m_1)$
	$m_{\text{out}} := (l_1 \wedge l_2) \oplus t$
	$l_{\text{out}} := m_1 \wedge m_2) \oplus t$
Subtraction	$t := (l_1 \wedge l_2) \oplus (m_2 \wedge m_1);$
	$m_{\text{out}} := (l_1 \wedge m_2) \oplus t;$
	$l_{\text{out}} := (m_1 \wedge l_2) \oplus t;$
Multiplication	$m_{\text{out}} := (l_1 \vee m_2) \wedge (m_1 \vee l_2);$
	$l_{\text{out}} := (l_1 \vee l_2) \wedge (m_1 \vee m_2);$
MUX	$m_{\text{out}} := (m_2 \vee s) \wedge (m_1 \vee (-s));$
	$l_{\text{out}}[i] := (l_2 \vee s) \wedge (l_1 \vee (-s));$

6.3.2 Integer packing and operations:

Another method that was explored for optimization was integer packing. At the last stage of our PRF, we multiply a Z_3 matrix R by a binary vector that is represented as a trinary vector. Since each number is a trinary number, the multiplication of each row by the output vector can be any number between -256 to 256. To explore this option, we represented each number by 9 bits, and

packed 7 numbers into each word. This was expected to result in time saving of up to $\times 7$. Testing this option indeed proved to be significantly slower than the first method of packing. We therefore continued using the first packing method of packing instead. Note: since each trinary number can be either 0,1 or 2, we can add up to 255 numbers and never exceed 511. Moreover, if we take a random sample, we can add 256 numbers and the probability of carryover is negligible. We use this to multiply a packed trinary vector with a ternary matrix with 256 columns (see algorithm below).

$x_0 \ldots x_6 \epsilon Z_3 = 0, 1, 2$

$\rightarrow x = \sum_{i=0}^{6} (512^i) \times c_i, x_i \epsilon [0...(2^{(}64) - 1)]$

6.3.3 Matrix multiplication - using a lookup table:

Phase 2 of the protocols (WPRF, OPRF) outputs m bits $\in \mathbb{Z}_2$. Phase 3 takes output of phase 2 and performs a matrix vector multiplication with a public randomization matrix $R \in \mathbb{Z}_3^{t \times m}$ to output $\mathbb{Z}_3^t$.

Naive method of performing this matrix-vector multiplication will be to multiply and add individual bits, which will take a long time. An optimized and faster version was also implemented as a function using packed bits. This packed version of multiplication runs way faster than the naive implementation. In order to further optimize the run-time, a lookup table is used to perform the task.

Lookup table is a matrix whose elements are precomputed product value of packed inputs(output of phase 2) with all possible combinations of similar sized packed values. Suppose, if inputs are packed as a 8-bit word, the matrix will contain all possible combination of product of 8-bit input and all values from 0 to $2^8 - 1$.

Implementing protocol using the lookup method consists of two sub-stages in phase 3:

- **Preprocessing Stage:** Creation of lookup table

- **Multiplication Stage:** Usage of lookup table

The prime advantage of replacing matrix vector multiplication with lookup table is speed improvement. This is due to creation of lookup table at the preprocessing stage of the protocol. While creation of a lookup table takes a prolong period of time, the multiplication stage is as simple as accessing an element from a matrix. This speeds up the last stage of the protocol, hence significantly improving the overall speed of execution.

To this end, the randomization matrix is divided into l matrices of size $81 \times (256/l)$. A lookup table of size $l \times 2^{(256/l)}$ is created during the pre-processing stage. Using lookup table results in a significantly larger data storage (as well as increased preprocessing time). Therefore, our goal was to find the number of sub-tables that should be used to provide faster protocol run-time while still maintaining a reasonable storage space.

To this end, we first use $l = 16$, therefore dividing the randomization matrix into 16 matrices of size 81×16. A lookup table of size 16×2^{16} is created during the pre-processing stage. During runtime, the input is divided to MSB and LSB, and each consecutive 16 bits (16 rows in the matrix) are used as separate input to the lookup table. Similarly, we explored dividing the table into 8 matrices of size 81×32, resulting in a lookup table of size 8×2^{32}. However, due to the large table size, this crashed the system during runtime and we were not able to get results for it.

6.4 Putting it all together

- **Our newly Implemented WPRF protocol was 50.7% faster and OPRF was 54% than implemented WPRF protocol described in TCC'18 [1]**

- **Using lookup table, the speed increase was significant at 70% and 76.5% for WPRF and WPRF respectively.**

7 Analysis

Table 2 includes the computation and communication results for the different protocol.

Protocol	Round Complexity	Online Communication	Preprocessing Size	Computation Operation
Centralized WPRF	-	-	-	9K
Distributed WPRF [1]	4	5	2K	45K
Our)WPRF	3	5	1M	13K
Our OPRF	3	5	1M	11K

Table 2: Communication Analysis of different protocols. The protocols were optimized using both the bit packing and the lookup table

8 Benchmarking

We compare our run-time to discrete-log based PRFs. To this end, we use the lib sodium library [2]. The library uses elliptic curve 252 bits, and includes a function that performs scalar multiplication ('crypto_scalarmult_ed25519').

9 Experimental Results

The system was tested using Ubuntu Server 18.04 on a t2.medium AWS environment. To record the timings, the code was run in a loop 1000 times. Below are run-time results for running a single instance of PRF in microsecond. The results include both centralized and distributed versions of the PRF. In order to increase efficiency, packing and lookup tables were used. The packing indicates both the Z_2 and Z_3 packing.

Protocol	Packed	lookup	Runs/sec	Runtime(μ sec)
Centralized weak PRF	N	N	50K	20.2
Centralized weak PRF	Y	N	65.4K	18.5
Centralized weak PRF	Y	Y	165K	6.08

Table 3: Comparison of the run-time and computation of the different Optimized Centralized wPRF.

Protocol	Runs/sec	Runtime(μ sec)
Fully Distributed WPRF[1]	24K	40.56
Fully Distributed WPRF	82K	12.12
OPRF with lookup	104K	9.52
Discrete log-based PRF	12K	**86.07**

Table 4: Optimized protocols runtime. These protocols utilize both the bit-packing and lookup table optimization

10 Conclusion and Future Work

11 Acknowledgement

12 References

References

[1] Dan Boneh et al. Exploring Crypto Dark Matter: New Simple PRF Candidates and Their Applications. Cryptology ePrint Archive, Report 2018/1218. `https://eprint.iacr.org/2018/1218`. 2018.

[2] libsodium 1.0.18-stable. `https://libsodium.gitbook.io/doc/`. Online; December 31 2020. 2020.