Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 13 (2013)

One-Round Multi-Party Communication Complexity of Distinguishing Sums

Daniel Apon* Jonathan Katz* Alex J. Malozemoff*

Abstract

We consider an instance of the following problem: Parties $P_1, \ldots, P_k$ each receive an input x_i, and a coordinator (distinct from each of these parties) wishes to compute $f(x_1, \ldots, x_k)$ for some predicate f. We are interested in *one-round* protocols where each party sends a single message to the coordinator; there is no communication between the parties themselves. What is the minimum communication complexity needed to compute f, possibly with bounded error?

We prove tight bounds on the one-round communication complexity when f corresponds to the promise problem of *distinguishing sums* (namely, determining which of two possible values the $\{x_i\}$ sum to) or the problem of determining whether the $\{x_i\}$ sum to a particular value. Similar problems were studied previously by Nisan and in concurrent work by Viola. Our proofs rely on basic theorems from additive combinatorics, but are otherwise elementary.

1 Introduction

Consider the following general problem: There are k parties $P_1, \ldots, P_k$, with each party P_i holding input x_i. A central coordinator (distinct from each of the parties) wants to learn $f(x_1, \ldots, x_k)$ for some fixed boolean function (or partial function) f. We are interested in *one-round* protocols where each party sends a single message to the coordinator and the coordinator then computes the result; there is no communication between the parties, nor does the coordinator send anything to the parties. A trivial solution, of course, is for each party P_i to send x_i to the coordinator, who then applies f to the complete set of inputs and thus obtains the correct result. For which functions f can the total communication complexity be reduced, possibly with bounded error?

Let $\mathbb{G}$ denote an abelian group and assume each party's input lies in $\mathbb{G}$. We study the communication complexity of two (related) functions in the model described above.

Definition 1. *Fix distinct $g_0, g_1 \in \mathbb{G}$. The k-party* Sum-Distinguish *problem (relative to g_0, g_1) is defined by letting f be the partial function given by*

$$f(x_1, \ldots, x_k) = \begin{cases} 1 & \text{if } \sum_i x_i = g_1 \\ 0 & \text{if } \sum_i x_i = g_0 \end{cases}.$$

Definition 2. *Fix $g \in \mathbb{G}$. The k-party* Sum-Equal *problem (relative to g) is defined by letting f be the function given by*

$$f(x_1, \ldots, x_k) = \begin{cases} 1 & \text{if } \sum_i x_i = g \\ 0 & \text{otherwise} \end{cases}.$$

*Dept. of Computer Science, University of Maryland. Email: {dapon,jkatz,amaloz}@cs.umd.edu.

ISSN 1433-8092

We explore the communication complexity of solving the above for $\mathbb{G} = \mathbb{Z}_p$ (p prime) and $\mathbb{G} = \mathbb{Z}$.[1] Our proofs rely on the generalized Cauchy-Davenport Theorem, but otherwise use only elementary arguments. Our results can be summarized as follows:

- For Sum-Distinguish with $\mathbb{G} = \mathbb{Z}$ or $\mathbb{G} = \mathbb{Z}_p$, p prime, we show a deterministic protocol with total communication complexity $k \log k + O(k)$; note that the communication in the latter case is independent of p. For $\mathbb{G} = \mathbb{Z}_p$, we prove a lower bound of $k \cdot \min\{\log k, \log p\} - k$ on the communication complexity of any deterministic protocol.
- For Sum-Equal with $\mathbb{G} = \mathbb{Z}$ or $\mathbb{G} = \mathbb{Z}_p$, p prime, we show a protocol using public randomness with error ϵ and total communication complexity $k \log k/\epsilon + O(k)$. A lower bound (for deterministic protocols and $\mathbb{G} = \mathbb{Z}_p$) is implied by our lower bound for Sum-Distinguish.

We also briefly consider the case $\mathbb{G} = \mathbb{Z}_N$ for square-free N.

1.1 Motivation

The problems above are natural in the *number-in-hand* model of multi-party communication complexity, and variants of the Sum-Distinguish and Sum-Equal problems have been considered in prior work [8, 7, 4, 5, 10], sometimes for $k = 2$ only. (We survey prior results in the next section.)

Our motivation, though, comes from the domain of *distributed intrusion detection*. The goal of distributed intrusion-detection systems (DIDS) is to monitor a network across a number of hosts in order to detect aberrant behavior (indicating a potential intrusion) and, if detected, raise an alarm. In typical operation of DIDS, each host records some observations over a specified time period; at the end of this period, each of those hosts sends all the data it has recorded to a central coordinator, which then determines — based on the aggregate data from all the hosts — whether or not to issue an alarm. In some systems (e.g., when the hosts are geographically distributed, when communication is over a low-bandwidth channel, and/or when the volume of data recorded at each host is huge), reducing the communication becomes critical. While there has been some work aimed at reducing the communication complexity of DIDS [2, 9, 6], we are not aware of any prior theoretical study of the problem.

If we model the decision of the coordinator by some predicate f computed over the data $x_1, \ldots, x_k$ recorded by each host, we recover exactly the general problem being considered here. (For the application to distributed intrusion detection, direct communication between the hosts would typically be impossible, and it would be undesirable for the coordinator to have to send data to the hosts.) While Sum-Distinguish and Sum-Equal are too simplistic to capture real-world decision procedures, they were chosen to correspond to the "DIDS-like" problems of distinguishing between a "good" system state g_0 and a "bad" system state g_1 (in the case of Sum-Distinguish), or identifying when the system is in one particular "bad" state g (in the case of Sum-Equal).

1.2 Prior Work

For the case of Sum-Equal with $\mathbb{G} = \mathbb{Z}$ and where each party's input is an n-bit integer, Nisan [7] shows a randomized protocol with total communication complexity $O(k \log n)$. Our deterministic protocol achieves better communication complexity $k \log k + O(k)$ when $k < n$. In concurrent and

[1] In the first case, each party's input is an arbitrary element of $\mathbb{Z}_p$; in the second case, each party's input is an n-bit integer, with n being an additional parameter of the problem.

independent work, Viola [10] studies SUM-EQUAL with $\mathbb{G} = \mathbb{Z}_p$, and shows $\Theta(k \log k)$ upper and lower bounds on the communication complexity for certain ranges of k and p. Our protocols for SUM-EQUAL achieve similar bounds for more general k, p, and using different tools.

Our protocols use a direct, combinatorial perspective that (along the way) explores a new connection between communication complexity and additive combinatorics that may be appealing in its own right. It will be interesting to explore other connections between these fields.

1.3 Organization

In Section 2, we recall the necessary preliminaries from additive combinatorics. In Section 3, we prove upper and lower bounds for the SUM-DISTINGUISH problem over $\mathbb{Z}_p$. In Section 4, we give a randomized protocol for SUM-EQUAL over $\mathbb{Z}_p$. In Section 5, we show how our protocols can be extended to work over $\mathbb{Z}$ or $\mathbb{Z}_N$ for m the product of few primes.

2 Preliminaries

We let $\mathbb{G}$ denote an abelian group, written additively. $\mathbb{Z}$ denotes the integers, and $\mathbb{Z}_p$ is the group $\{0, \ldots, p-1\}$ under addition modulo p. We use "log" to refer to logarithms base 2.

2.1 Tools from Additive Combinatorics

We utilize two well-studied, fundamental objects from additive combinatorics: *sumsets* and *arithmetic progressions*.

Definition 3. *For (not necessarily distinct) sets $A_1, \ldots, A_k \subseteq \mathbb{G}$, define their* sumset *as $\sum_{i=1}^k A_i = A_1 + \cdots + A_k \stackrel{\text{def}}{=} \left\{ \sum_{i=1}^k a_i \mid a_1 \in A_1, \ldots, a_k \in A_k \right\}$. That is, $\sum_i A_i$ is the set of all possible sums obtainable by choosing one element from each set A_i.*

In our constructions we use sumsets of arithmetic progressions, i.e. sequences of integers with common difference D. We refer to these as *D-APs*.

Definition 4. *Fix a prime p and a difference $D \neq 0 \bmod p$. For any $b \in \{0, \ldots, D-1\}$, let*

$$A_{(b)} \stackrel{\text{def}}{=} \left\{ b,\ b+D,\ b+2D,\ \ldots,\ b + \left(\left\lfloor \frac{p-1-b}{D} \right\rfloor \right) D \right\} \subseteq \mathbb{Z}_p$$

*denote the D-*AP *in $\mathbb{Z}_p$* with base *b.*

Note that we only consider D-APs of maximal size with no "wrap-around"; i.e., the base b is less than D, and the progression contains $b + iD$ for all $i \geq 0$ with $b + iD < p$. As an example, the maximal 7-APs in $\mathbb{Z}_{19}$ are

$$A_{(0)} = \{0, 7, 14\}; \quad A_{(1)} = \{1, 8, 15\} \quad A_{(2)} = \{2, 9, 16\}$$
$$A_{(3)} = \{3, 10, 17\}; \quad A_{(4)} = \{4, 11, 18\}; \quad A_{(5)} = \{5, 12\}; \quad A_{(6)} = \{6, 13\}.$$

For our lower bounds, we use the generalized Cauchy-Davenport Theorem [1, 3].

Theorem 5 (Generalized Cauchy-Davenport Theorem)**.** *For a prime p, and k (not necessarily distinct) nonempty sets $A_1, \ldots, A_k \subseteq \mathbb{Z}_p$,*

$$\left|\sum_{i=1}^{k} A_i\right| \geq \min\left\{p,\ \sum_{i=1}^{k} |A_i| - k + 1\right\}.$$

For our constructions, we rely on the fact that sumsets of D-APs achieve the above minimum.

Lemma 6. *For a prime p, and k (not necessarily distinct) D-APs $A_1, \ldots, A_k \subseteq \mathbb{Z}_p$,*

$$\left|\sum_{i=1}^{k} A_i\right| = \min\left\{p,\ \sum_{i=1}^{k} |A_i| - k + 1\right\}.$$

Proof. We prove the lemma for $k = 2$; the general case follows by induction. Let A, B denote the two sets in question. By Theorem 5, we have $|A + B| \geq \min\{p, |A| + |B| - 1\}$. It remains to upper bound $|A + B|$.

Write $A = \{b_A + iD \mid 0 \leq i < |A|\}$, $B = \{b_B + i'D \mid 0 \leq i' < |B|\}$ with $0 \leq b_A, b_B < D$. Then

$$\begin{aligned} A + B &= \left\{b_A + b_B + Di + Di' \bmod p \mid 0 \leq i < |A|,\ 0 \leq i' < |B|\right\} \\ &= \left\{b_A + b_B + Di'' \bmod p \mid 0 \leq i'' \leq |A| + |B| - 2\right\}. \end{aligned}$$

So $|A + B|$ contains at most $|A| + |B| - 1$ elements, giving the desired bound. □

2.2 Notions of Distance and Contiguity

In Section 3.1, we use a notion of distance between two elements $g_0, g_1 \in \mathbb{Z}_p$. Specifically, we define their distance relative to some difference D to be the minimum number of additions or subtractions by D (modulo p) needed to map g_0 to g_1. We define this formally next.

Definition 7. *Fix a prime p and a difference $D \neq 0 \bmod p$. For any $g_0, g_1 \in \mathbb{Z}_p$, define the* distance from g_0 to g_1 *(relative to p, D) as*

$$\mathsf{dist}_{p,D}(g_0, g_1) \stackrel{\text{def}}{=} \min\left\{(g_1 - g_0)D^{-1} \bmod p,\ (g_0 - g_1)D^{-1} \bmod p\right\}.$$

(The minimum is taken by viewing each term as an integer in $\{0, \ldots, p-1\}$.)

We say $g_0, g_1 \in \mathbb{Z}_p$ are *adjacent* (with respect to $\mathsf{dist}_{p,D}$) if $\mathsf{dist}_{p,D}(g_0, g_1) = 1$. We say a set $A \subseteq \mathbb{Z}_p$ is *contiguous* (relative to D) if it can be ordered so that all adjacent elements in the ordering are adjacent with respect to $\mathsf{dist}_{p,D}$. When $|A| = 1$, A is vacuously contiguous. Clearly D-APs are contiguous; we observe that sumsets of D-APs are also contiguous.

Lemma 8. *Fix a prime p, difference $D \neq 0 \bmod p$, and any D-APs $A_1, \ldots, A_k \subseteq \mathbb{Z}_p$. Then $\sum_{i=1}^{k} A_i$ is contiguous relative to D.*

Proof. From the proof of Lemma 6, for any D-APs A and B we have

$$A + B = \{b_A + b_B + iD \bmod p \mid 0 \leq i \leq |A| + |B| - 2\},$$

which is contiguous by definition. Induction on k completes the proof. □

Corollary 9. *Fix a prime p, difference $D \neq 0 \bmod p$, and D-APs $A_1, \ldots, A_k \subseteq \mathbb{Z}_p$. For any $g_0, g_1 \in \mathbb{Z}_p$, if $\mathsf{dist}_{p,D}(g_0, g_1) \geq |\sum_{i=1}^{k} A_i|$ then g_0 and g_1 cannot both be in $\sum_{i=1}^{k} A_i$.*

Consider again the example of $\mathbb{Z}_{19}$ with $D = 7$. Then $A_{(2)} + A_{(3)} = \{2, 9, 16\} + \{3, 10, 17\} = \{5, 12, 0, 7, 14\}$ is contiguous, and $|A_{(2)} + A_{(3)}| = 5$. Taking $2, 5 \in \mathbb{Z}_{19}$, we have $\mathsf{dist}_{19,7}(5, 2) = 5 \geq |A_{(2)} + A_{(3)}|$ and, indeed, $5 \in A_{(2)} + A_{(3)}$ but $2 \notin A_{(2)} + A_{(3)}$.

3 Sum-Distinguish over $\mathbb{Z}_p$

3.1 A Deterministic Protocol

Corollary 9 suggests a technique for efficiently distinguishing two sums. Say k parties wish to determine whether their inputs $x_1, \ldots, x_k$ sum to g_0 or g_1 (modulo p). For some fixed, agreed-upon difference D (we discuss how to set D below), each party P_i sends to the coordinator the *index* of the D-AP A_i in which its input x_i lies. The coordinator thus learns that the sum $\sum_i x_i$ lies in the sumset $A \stackrel{\text{def}}{=} \sum_{i=1}^{k} A_i$. As long $\mathsf{dist}_{p,D}(g_0, g_1) \geq |A|$, it cannot be the case that both g_0 and g_1 are in A; in that case, the coordinator learns the sum by checking which of g_0, g_1 lies in A.

The main difficulty in implementing the above is that g_0, g_1 may be very "close." In that case, in order to ensure that the above succeeds we need to ensure that $|A|$ is small. This, in turn, requires the D-APs to be small, which means that there are more of them. Since the communication from each party is the logarithm of the number of D-APs, this makes the communication complexity worse. Ideally, we would like to set D independently of the relative distance between g_0 and g_1.

A solution is to have the parties "shift" their inputs by each locally multiplying them (modulo p) by an agreed-upon constant c. The problem then reduces to distinguishing whether the shifted inputs sum to $g_0' \stackrel{\text{def}}{=} c \cdot g_0 \bmod p$ or $g_1' \stackrel{\text{def}}{=} c \cdot g_1 \bmod p$. The insight is that regardless of g_0, g_1, we can set c appropriately to ensure that g_0' and g_1' are "far apart."

We proceed with the details, beginning with some preliminary lemmas.

Lemma 10. *Fix a prime $p > 2$ and a difference $D \neq 0 \bmod p$. Then for any distinct $g_0, g_1 \in \mathbb{Z}_p$ there exists a value $c \neq 0 \bmod p$ such that $\mathsf{dist}_{p,D}(c \cdot g_0 \bmod p,\ c \cdot g_1 \bmod p) = \frac{(p-1)}{2}$.*

Proof. Set $c = \frac{(p-1)}{2} D (g_1 - g_0)^{-1} \bmod p$. Then

$$\begin{aligned} c(g_1 - g_0)D^{-1} &= \frac{(p-1)}{2} D (g_1 - g_0)^{-1} (g_1 - g_0) D^{-1} \bmod p \\ &= \frac{(p-1)}{2} \bmod p. \end{aligned}$$

Since $(p-1)/2 < -(p-1)/2 \bmod p$ (viewing the right-hand term as an integer in $\{1, \ldots, p-1\}$), this completes the proof. □

Lemma 11. *Fix a prime $p > 5$, and integer $k < p/4$. Set $D = \left\lceil \frac{2kp}{(p-3)} \right\rceil < p$. Then for any D-APs $A_1, \ldots, A_k \subseteq \mathbb{Z}_p$, we have $\left| \sum_{i=1}^{k} A_i \right| \leq \frac{(p-1)}{2}$.*

Proof. By Lemma 6,

$$\left| \sum_{i=1}^{k} A_i \right| = \min \left\{ p,\ \sum_{i=1}^{k} |A_i| - k + 1 \right\}.$$

Observe that $|A_i| \leq \lceil \frac{p}{D} \rceil \leq \frac{p}{D} + 1$. Therefore,

$$\begin{aligned} \sum_{i=1}^{k} |A_i| - k + 1 &\leq \sum_{i=1}^{k} \left(\frac{p}{D} + 1 \right) - k + 1 \\ &\leq \frac{kp(p-3)}{2kp} + k - k + 1 = \frac{(p-1)}{2}, \end{aligned}$$

completing the proof. □

Theorem 12. *There is a universal constant C such that for any prime p and positive integer k there is a k-party, one-round, deterministic protocol for* Sum-Distinguish *over $\mathbb{Z}_p$ having communication complexity $k \log k + C \cdot k$.*

Proof. There is a trivial protocol having communication complexity $k \cdot \lceil \log p \rceil$, so the theorem is trivially true if $p \leq 5$ or $k \geq p/4$. In what follows we therefore assume $p > 5$ and $k < p/4$.

Fix arbitrary, distinct $g_0, g_1 \in \mathbb{Z}_p$. The protocol for solving Sum-Distinguish relative to g_0, g_1 is as follows. Set c as in Lemma 10, and D as in Lemma 11. Party P_i, holding input x_i, computes

$$b_i = ((c \cdot x_i) \bmod p) \bmod D$$

and sends b_i to the coordinator. (Note that b_i is the base for the D-AP $A_{(b_i)}$ containing $c \cdot x_i \bmod p$.) The coordinator outputs 0 if $c \cdot g_0 \bmod p$ is in $\sum_{i=1}^{k} A_{(b_i)}$, and outputs 1 otherwise.

If $\sum_i x_i = g_0 \bmod p$ then $\sum_i c \cdot x_i = c \cdot g_0 \bmod p$, and it is immediate that the coordinator outputs the correct answer 0. So, assume instead that $\sum_i x_i = g_1 \bmod p$. Then $c \cdot g_1 \bmod p$ is in $\sum_i A_{(b_i)}$. Since $\mathsf{dist}_{p,D}(c \cdot g_0 \bmod p, c \cdot g_1 \bmod p) = (p-1)/2$ (by Lemma 10) and $\left|\sum_i A_{(b_i)}\right| \leq (p-1)/2$ (by Lemma 11), we conclude from Corollary 9 that $c \cdot g_0 \bmod p$ is not in $\sum_i A_{(b_i)}$. Hence, in this case the coordinator outputs the correct answer 1.

The communication complexity is exactly $k \cdot \lceil \log D \rceil$ bits. Since $D \leq \frac{2kp}{p-3} + 1 \leq C' \cdot k$ for some constant C' independent of p and k, this completes the proof. □

Efficient implementation. We note that the coordinator can be implemented to run efficiently. (It is clear that the parties can run efficiently.) First note that from b_i the coordinator can efficiently compute $\left|A_{(b_i)}\right| = \lceil \frac{p-1-b_i}{D} \rceil + 1$. It can then compute $d \stackrel{\text{def}}{=} \left|\sum_i A_{(b_i)}\right| = \sum_{i=1}^{k} |A_{(b_i)}| - k + 1$. Finally, the coordinator can check whether $c \cdot g_0 \in \sum_i A_{(b_i)}$ by computing $b^* = \sum_i b_i \bmod p$ and then checking whether $(c \cdot g_0 - b^*) \cdot D^{-1} \bmod p$ is less than d.

3.2 A Lower Bound for Deterministic Protocols

In the following, we consider one-round protocols in which each party always sends exactly t bits to the coordinator, for some t. We say any such protocol has *per-party communication complexity* t.

The basic idea of the lower bound is as follows. Each message $m \in \{0,1\}^t$ from party P_1, say, defines a set $A_{1,m}$ of possible inputs x_1 (namely, those inputs on which P_1 would send m). Given the messages $m_1, \ldots, m_k$ sent by all the parties, the coordinator learns only that the sum $\sum_i x_i$ lies in the sumset $\sum_i A_{i,m_i}$. If we can show that there exist some $m_1, \ldots, m_k$ for which $\sum_i A_{i,m_i}$ contains both g_0 and g_1, then there must be some set of inputs on which the protocol outputs the wrong result. The crux of the proof is to show that if t is too small, then there exist $m_1, \ldots, m_k$ for which $\sum_i A_{i,m_i} = \mathbb{Z}_p$, and hence the sumset does indeed contain both g_0 and g_1.

Theorem 13. *Fix prime p and positive integer $k > 1$. If $t \leq \min\{\log((k-1)/2), \log(p/2)\}$, there is no deterministic, k-party protocol for* Sum-Distinguish *over $\mathbb{Z}_p$ (relative to* any *$g_0, g_1 \in \mathbb{Z}_p$) with per-party communication complexity t.*

Proof. Fix some deterministic protocol for Sum-Distinguish over $\mathbb{Z}_p$ (relative to some $g_0, g_1 \in \mathbb{Z}_p$) with per-party communication complexity t. The protocol defines for each party P_i a partition $A_{i,1}, \ldots, A_{i,2^t}$ of $\mathbb{Z}_p$, where $A_{i,j}$ is the set of inputs which cause P_i to send j to the coordinator.

For each party P_i there exists an m_i such that $|A_{i,m_i}| \geq p/2^t$. Moreover, there is a legal set of inputs for the parties such that $P_1, \ldots, P_{k-1}$ send $m_1, \ldots, m_{k-1}$, respectively. (Simply take $x_i \in A_{i,m_i}$ for $i = 1, \ldots, k-1$, and then let $x_k \in \{g_0 - \sum_{i=1}^{k-1} x_i,\ g_1 - \sum_{i=1}^{k-1} x_i\}$.) When the coordinator receives $m_1, \ldots, m_{k-1}$ then, even if it is additionally given P_k's input x_k, the coordinator learns only that the sum $\sum_i x_i$ of the parties' inputs lies in the set $x_k + \sum_{i=1}^{k-1} A_{i,m_i}$. By Theorem 5, however, we have

$$\left|\sum_{i=1}^{k-1} A_{i,m_i}\right| \geq \min\left\{p,\ \sum_{i=1}^{k-1} \frac{p}{2^t} - (k-1) + 1\right\} = \min\left\{p, \frac{(k-1)p}{2^t} - k + 2\right\}.$$

If $p \geq k-1$ then $t \leq \log((k-1)/2)$ and

$$\frac{(k-1)p}{2^t} - k + 2 \geq 2p - k + 2 > p.$$

On the other hand, if $k - 1 > p$ then $t \leq \log(p/2)$ and

$$\frac{(k-1)p}{2^t} - k + 2 \geq k > p.$$

In either case, then, we must have $\left|\sum_{i=1}^{k-1} A_{i,m_i}\right| \geq p$ and so $\sum_{i=1}^{k-1} A_{i,m_i} = \mathbb{Z}_p$. This implies that there exist inputs $x_1, x_1' \in A_{1,m_1}, \ldots, x_{k-1}, x_{k-1}' \in A_{k,m_{k-1}}$ with $x_k + \sum_{i=1}^{k-1} x_i = g_0$ and $x_k + \sum_{i=1}^{k-1} x_i' = g_1$. But then there exists some set of legal inputs for the parties on which the coordinator outputs an incorrect result. □

4 A Randomized Protocol for Sum-Equal over $\mathbb{Z}_p$

Our protocol for Sum-Equal is similar to our protocol for Sum-Distinguish. Namely, each party P_i scales its input x_i by some value c and sends the index of the D-AP A_i that contains the scaled value $c \cdot x_i \bmod p$; the coordinator outputs 1 iff $c \cdot g \in \sum_i A_i$.

Note that the coordinator never errs if $\sum_i x_i = g$, and so we need only analyze the case when $\sum_i x_i \neq g$. In the case of Sum-Distinguish, we are guaranteed that $\sum_i x_i \in \{g_0, g_1\}$ and so we set c to some fixed value such that cg_0 and cg_1 are "far apart." The problem here is that $\sum_i x_i$ can be arbitrary. To deal with this, we have the parties select $c \in \mathbb{Z}_p$ uniformly at random using the public randomness. If $\sum_i x_i = g' \neq g$ then the protocol will succeed as long as cg and cg' are sufficiently "far apart" as before. By setting the parameters of the protocol appropriately, we ensure that this happens with high probability over choice of c.

Lemma 14. *Fix a prime $p > 2$, a difference $D \neq 0 \bmod p$, and $\xi \in (0,1)$. Then for any distinct $g, g' \in \mathbb{Z}_p$ there are at least $\xi \cdot (p-1)$ values $c \neq 0 \bmod p$ such that*

$$\mathsf{dist}_{p,D}(c \cdot g \bmod p,\ c \cdot g' \bmod p) > (1-\xi) \cdot \frac{(p-1)}{2}.$$

Proof. Set $\delta \stackrel{\text{def}}{=} \lceil \xi \cdot (p-1)/2 \rceil$, and take any d in the set $\left\{\frac{(p-1)}{2} - \delta + 1, \ldots, \frac{(p-1)}{2} + \delta\right\}$ of size 2δ. Set $c = d \cdot D(g-g')^{-1} \bmod p$. Then

$$c(g-g')D^{-1} = d \cdot D(g-g')^{-1}(g-g')D^{-1} = d \bmod p.$$

So,

$$\begin{aligned}\mathsf{dist}_{p,D}(c\cdot g \bmod p,\ c\cdot g' \bmod p) &= \min\{c(g-g')D^{-1} \bmod p,\ c(g'-g)D^{-1} \bmod p\} \\ &= \min\{d, p-d\} \ \geq\ \frac{(p-1)}{2} - \delta + 1,\end{aligned}$$

completing the proof. □

Lemma 15. *Fix a prime $p > 5$, an integer $k < p/4$, and $\epsilon > \frac{2k}{p-3}$. Set $D = \left\lceil \frac{2kp}{\epsilon(p-3)} \right\rceil < p$. Then for any D-APs $A_1, \ldots, A_k \subseteq \mathbb{Z}_p$, we have $\left|\sum_{i=1}^k A_i\right| < \epsilon \cdot \frac{(p-1)}{2} + 1$.*

Proof. By Lemma 6,

$$\left|\sum_{i=1}^k A_i\right| = \min\left\{p,\ \sum_{i=1}^k |A_i| - k + 1\right\}.$$

Observe that $|A_i| \leq \lceil \frac{p}{D} \rceil \leq \frac{p}{D} + 1$. Therefore,

$$\begin{aligned}\sum_{i=1}^k |A_i| - k + 1 &\leq \sum_{i=1}^k \left(\frac{p}{D} + 1\right) - k + 1 \\ &\leq \frac{kp\epsilon(p-3)}{2kp} + k - k + 1 < \frac{\epsilon(p-1)}{2} + 1,\end{aligned}$$

completing the proof. □

Theorem 16. *There is a universal constant C such that for any prime p, positive integer k, and $\epsilon \in (0,1)$, there is a k-party, one-round protocol for* Sum-Equal *over $\mathbb{Z}_p$ using public randomness, with error at most ϵ and communication complexity $k \log k/\epsilon + C \cdot k$.*

Proof. There is a trivial protocol with communication complexity $k \cdot \lceil \log p \rceil$, so the theorem is true if $p \leq 5$ or $k \geq p/4$ or $\epsilon \leq \frac{2k}{p-3}$. In what follows we therefore assume $p > 5$, $k < p/4$, and $\epsilon > \frac{2k}{p-3}$.

The protocol for solving Sum-Equal is as follows. Set D as in Lemma 15, and use the public randomness to choose uniform $c \in \mathbb{Z}_p \setminus \{0\}$. Party P_i, holding input x_i, computes

$$b_i = ((c \cdot x_i) \bmod p) \bmod D$$

and sends b_i to the coordinator. The coordinator outputs 1 if $c \cdot g \bmod p$ is in $\sum_{i=1}^k A_{(b_i)}$, and outputs 1 otherwise.

If $\sum_i x_i = g \bmod p$ then $\sum_i c \cdot x_i = c \cdot g \bmod p$ and the coordinator always outputs the correct answer 1. Now say $\sum_i x_i = g' \neq g \bmod p$. Then $c \cdot g' \bmod p$ is in $\sum_i A_{(b_i)}$. Using Lemma 15, we have $\left|\sum_i A_{(b_i)}\right| < \epsilon(p-1)/2 + 1$. Using Lemma 14, with probability at least $\xi \stackrel{\text{def}}{=} 1 - \epsilon$ we have $\mathsf{dist}_{p,D}(c \cdot g \bmod p,\ c \cdot g' \bmod p) > \epsilon(p-1)/2$. Assuming that to be the case, we have

$$\mathsf{dist}_{p,D}(c \cdot g \bmod p,\ c \cdot g' \bmod p) \geq \left|\textstyle\sum_i A_{(b_i)}\right|$$

(note that both sides of the above are integers), and so we conclude from Corollary 9 that $c \cdot g \bmod p$ is not in $\sum_i A_{(b_i)}$. We thus see that with probability at least $1 - \epsilon$ the coordinator outputs the correct answer 0.

The communication complexity is exactly $k \cdot \lceil \log D \rceil$ bits. Since $D \leq \frac{2kp}{\epsilon(p-3)} + 1 \leq C' \cdot k/\epsilon$ for some constant C' independent of p, k, and ϵ, this completes the proof. □

5 Protocols Over $\mathbb{Z}$ and $\mathbb{Z}_N$

In what follows, we show how to modify our protocols to work over the integers and in $\mathbb{Z}_N$ for square-free N.

Protocol over $\mathbb{Z}$. Working over $\mathbb{Z}$ is relatively easy. The parties are given inputs in $\{0, \ldots, 2^n - 1\}$. The maximum sum of all the inputs is $k2^n$, and the "target values" are at most that also. The parties choose the smallest prime $p > k2^n$, treat their inputs as lying in $\mathbb{Z}_p$, and run the protocol for $\mathbb{Z}_p$. Note that $\sum_i x_i = g$ over the integers iff $\sum_i x_i = g \bmod p$ by our choice of p.

Protocol over $\mathbb{Z}_N$. Assume N is square-free, and let $N = \prod_{i=1}^{m} p_i$ be the prime factorization of N. The parties can then work modulo each of the p_i, and rely on the Chinese remainder theorem for correctness. The complexity of the protocol scales with the number of prime factors m.

This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-11-2-0086. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] A. L. Cauchy. Recherches sur les nombres. *Journal de l'École Polytechnique*, 9:99–116, 1813.

[2] S. Cheung and A. Valdes. Malware characterization through alert pattern discovery. *Proc. 2nd USENIX Conference on Large-Scale Exploits and Emergent Threats (LEET)*, 2009.

[3] H. Davenport. On the addition of residue classes. *J. London Mathematical Society*, s1-10(1):30–32, 1935.

[4] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[5] P.B. Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. *J. Computer and System Sciences* 57(1):37–49, 1998.

[6] P. Ning, S. Jajodia, and X.S. Wang. Abstraction-based intrusion detection in distributed environments. *ACM Trans. Information and System Security* 4(4): 407–452, 2001.

[7] N. Nisan. The communication complexity of threshold gates. In *Combinatorics, Paul Erdős is Eighty (Volume 1)*, Bolyai Society Mathematical Studies, pp. 301–315, 1993.

[8] D.V. Smirnov. Shannon's information methods for lower bounds for probabilistic communication complexity. Master's thesis, Moscow University, 1988.

[9] A. Valdes and K. Skinner. Probabilistic alert correlation. *Proc. 4th International Symposium on Recent Advances in Intrusion Detection (RAID)*, 2001.

[10] E. Viola. The communication complexity of addition. *Proc. 24th Annual Symposium on Discrete Algorithms (SODA)*, 2013. Available at `http://eccc.hpi-web.de/report/2011/152`.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de