On the Structure of the Privacy Hierarchy

by

B. Chor, M. Geršb-Grauw, E. Kushilevitz

Technical Report #674
May 1991
On the Structure of the Privacy Hierarchy*

Benny Chor† Mihály Geréb-Grausz‡ Eyal Kushilevitz§

Abstract

We show that the privacy hierarchy for N-argument functions \( f(x_1, \ldots, x_N) \) which are defined over finite domains, has exactly \( \left\lceil \frac{N}{2} \right\rceil \) levels. We prove this by constructing, for any \( \left\lceil \frac{N}{2} \right\rceil \leq t \leq N - 2 \), an N-argument function which is \( t \) - private but not \( t + 1 \) - private.

1 Introduction

An N-argument function \( f(x_1, \ldots, x_N) \) is \( t \) - private if there exists a protocol for distributively computing \( f \), so that no coalition of \( \leq t \) parties can infer any additional information (in the information-theoretic sense) from the execution of the protocol (additional information is any information on inputs of non-coalition members which does not follow from inputs of coalition members and the value \( f(x_1, \ldots, x_N) \)). Ben-Or, Goldwasser and Wigderson [1] and Chaum, Crepeau and Damgard [5] have shown that over finite domains, every function can be computed \( \left\lceil \frac{N}{2} \right\rceil \) - privately. Some functions, like modular addition [3], are even \( N \) - private, while others, like Boolean OR, are \( \left\lceil \frac{N-1}{2} \right\rceil \) - private but not \( \left\lceil \frac{N}{2} \right\rceil \) - private [1].

For certain infinite families of functions (Boolean functions [7] and a certain class of symmetric functions [8]) it was proved that every function which is \( \left\lceil \frac{N}{2} \right\rceil \) - private is also \( N \)-private. No function which is \( t \) - private but not \( t + 1 \) - private, for \( \left\lceil \frac{N}{2} \right\rceil \leq t \leq N - 2 \), was known to date. We show that for every \( \left\lceil \frac{N}{2} \right\rceil \leq t \leq N - 2 \) there exists a function that is \( t \) - private but not \( t + 1 \) - private. This proves the existence of a dense privacy hierarchy.

2 Definitions and Background

We have a distributed network of \( N \) - synchronous, computationally unbounded parties \( P_1, P_2, \ldots, P_N \). Each pair of parties is connected by a secure (no eavesdropping) and reliable communication channel. At the beginning of an execution, each party \( P_i \) has an input \( a_i \) (no probability space is associated with the inputs). In addition, each party has a random input \( r_i \) taken from a

---

*Research supported by US-Israel Binational Science Foundation grant 88-00282.
†e-mail: benny@techsel.bitnet - Department of Computer Science, Technion, Haifa 32000, Israel.
‡ e-mail: gereb@cs.tufts.edu - Department of Computer Science, Tufts University, Medford, MA 02155.
§ e-mail: eyalk@techunix.bitnet - Department of Computer Science, Technion, Haifa 32000, Israel.

source of randomness \( R_i \) (the random inputs are independent). The parties wish to compute the value of a function \( f(x_1, x_2, \ldots, x_N) \). To this end, they exchange messages as prescribed by a (randomized) protocol \( \mathcal{F} \). Messages are sent in rounds, where in each round every processor can send a message to every other processor. Each message a party sends in the \( k \)-th round is determined using its input, its random input, the messages it received so far, and the identity of the receiver. As commonly assumed, the messages sent at each round are prefix free. The last message in the protocol contains the value of the function \( f(x_1, x_2, \ldots, x_N) \).

We say that a coalition (i.e. a set of parties) \( T \) does not learn any additional information (other than what follows from its input and the function value) from the execution of a randomized protocol \( \mathcal{F} \), which computes \( f \), if the following holds: For every two input vectors \( x, y \) that agree in their \( T \) entries (i.e. \( \forall i \in T: x_i = y_i \)) and satisfy \( f(x) = f(y) \), and for every choice of random inputs \( \{ r_i \}_{i \in T} \), the messages passed between \( T \) and \( T' \) are identically distributed. That is, for every communication \( S \),

\[
\Pr(S|x, \{ r_i \}_{i \in T}) = \Pr(S|y, \{ r_i \}_{i \in T}),
\]

where the probability space is over the random inputs of all parties in \( T \).

We say that a protocol \( \mathcal{F} \) for computing \( f \) is \( t \)-private if any coalition \( T \) of size \( \leq t \) does not learn any additional information from the execution of the protocol. We say that a function \( f \) is \( t \)-private if there exists a \( t \)-private protocol that computes it.

In the sequel, we will use two known lemmata. The first lemma states a necessary condition for \( t \)-privacy (\( t \geq \left\lfloor \frac{N}{2} \right\rfloor \)) of \( f \), in terms of \( 1 \)-privacy of a related two-argument function. The second lemma states a necessary condition for \( 1 \)-privacy of two-argument functions.

**Lemma 1:** \((\text{[7]})\) Let \( A_1, A_2, \ldots, A_N \) and \( B \) be non-empty sets, \( t \geq \left\lfloor \frac{N}{2} \right\rfloor \), and \( f: A_1 \times A_2 \times \ldots \times A_N \to B \) be \( t \)-private. Let \( S \subseteq \{1, 2, \ldots, N\} \) be any subset of size \( t \). Denote by \( D \) (resp. \( E \)) the Cartesian product of the \( A_i \) with \( i \in S \) (resp. \( i \in S \)). Then, viewing \( f \) as a two argument function \( f': D \times E \to B \), \( f' \) is \( 1 \)-private.

**Lemma 2:** \((\text{[7, 9]})\) Let \( D, E \) and \( B \) be non-empty sets, and \( f: D \times E \to B \) be \( 1 \)-private. For every \( d_1, d_2 \in D \), \( e_1, e_2 \in E \) and \( b \in B \), if \( f(d_1, e_1) = f(d_1, e_2) = f(d_2, e_1) = b \) then \( f(d_2, e_2) = b \).

### 3 The Hierarchy

**Theorem 1:** Let \( \left\lfloor \frac{N}{2} \right\rfloor \leq t \leq N - 2 \). There exists an \( N \)-argument function \( f_t \), which is \( t \)-private but not \( t + 1 \)-private.

**Proof:** For every \( t \) (\( \left\lfloor \frac{N}{2} \right\rfloor \leq t \leq N - 2 \)), let \( f_t : \{0, 1\}^N \to \{0, 1\}^N \cup \{0, 1\} \) be defined by

\[
\begin{align*}
f_t(0, 0, \ldots, 0, 0, x_{t+3}, \ldots, x_N) &= 0 \\
f_t(0, 0, \ldots, 0, 1, x_{t+3}, \ldots, x_N) &= 0 \\
f_t(1, 1, \ldots, 1, 0, x_{t+3}, \ldots, x_N) &= 0
\end{align*}
\]

2

Figure 1: \( g_t \) does not satisfy Lemma 2

\[
f_t(1,1,\ldots,1,1,x_{t+3},\ldots,x_N) = 1
\]

For all other inputs in \( \{0,1\}^N \) let \( f(\bar{x}) = \bar{0} \).

First we show that \( f_t \) is not \( t+1 \)-private. By Lemma 1 it is enough to demonstrate a partition \( S, S' \) of \( \{1,\ldots,N\} \) such that \( S \) is of size \( t+1 \), and the induced two-argument function is not \( 1 \)-private. We choose \( S = \{1,2,\ldots,t+1\} \). It is easy to verify (see Figure 1) that the induced two-argument function, \( g_t \), does not satisfy Lemma 2 and thus is not \( 1 \)-private.

Now, we prove that \( f_t \) is \( t \)-private. To do this we present an appropriate protocol, and show that it is \( t \)-private.

1. The party \( P_{t+2} \) shares its input, \( x_{t+2} \), among the parties \( P_1, P_2,\ldots,P_{t+1} \) using a secret-sharing scheme (for a definition see [10, 4]). This implies that \( P_1, P_2,\ldots,P_{t+1} \) together can reconstruct \( x_{t+2} \), while any subset of them does not have any information about \( x_{t+2} \).

2. The parties \( P_1, P_2,\ldots,P_{t+1} \) exchange their inputs.

3. If \( x_1 = x_2 = \ldots = x_{t+1} = 0 \) then the parties \( P_1, P_2,\ldots,P_{t+1} \) announce that the output is \( f_t(\bar{x}) = 0 \) and the protocol terminates.

4. If \( x_1 = x_2 = \ldots = x_{t+1} = 1 \) then the parties \( P_1, P_2,\ldots,P_{t+1} \) reconstruct \( x_{t+2} \), they announce that the output is \( f_t(\bar{x}) = x_{t+2} \), and the protocol terminates.

5. Otherwise, \( P_1, P_2,\ldots,P_{t+1} \) notify all the parties to exchange their inputs, and the output is \( f_t(\bar{x}) = \bar{z} \).

Now, we need to prove that the above protocol is indeed \( t \)-private. First, note that any output different from \( \{0,1\} \), completely determines the input vector and thus the privacy condition with respect to these vectors is satisfied by every protocol. Another simple case is where the output is 1. In this case the input vector is of the form

\[
(1,1,\ldots,1,1,x_{t+3},\ldots,x_N).
\]

However, in this case, the parties \( P_{t+3},\ldots,P_N \) do not take part in the protocol (except of receiving the final output) and therefore for each of the possible input vectors we have the same distribution of messages.
The interesting case is where the output is 0. Again, in this case, the parties $P_{t+3}, \ldots, P_N$ do not take part in the protocol. Therefore, they do not contribute any information to coalitions to which they belong. We discuss the possible coalitions (of size $\leq t$) according to several cases:

- Coalitions of size $\leq t$ which are subsets of $\{1, 2, \ldots, t+1\}$. Such a coalition should not be able to distinguish between input vectors of the form

$$ (0, 0, \ldots, 0, 0, z_{t+3}, \ldots, z_N) $$

and input vectors of the form

$$ (0, 0, \ldots, 0, 1, z_{t+3}, \ldots, z_N). $$

However, by the protocol, for such input vectors the parties $P_1, \ldots, P_{t+1}$ do not "open" the input $z_{t+2}$. Therefore, as $z_{t+2}$ is shared using a secret-sharing-scheme, every proper subset of $P_1, \ldots, P_{t+1}$ see the same distribution of messages in the case that $z_{t+2} = 0$ and in the case that $z_{t+2} = 1$. As all the other steps of the protocol are deterministic, then the distribution of messages seen by such a coalition, for any input vector of the above forms, is identical.

- The coalition $\{P_{t+2}\}$. In this case, $P_{t+2}$ should not be able to distinguish between input vectors of the form

$$ (0, 0, \ldots, 0, 0, z_{t+3}, \ldots, z_N) $$

and input vectors of the form

$$ (1, 1, \ldots, 1, 0, z_{t+3}, \ldots, z_N). $$

However, note that $P_{t+2}$ does not receive any message during the execution of the protocol (except for the output of the function). Therefore, the distribution of messages that it receives/sends on all these input vectors is identical.

- Coalitions that consist of $P_{t+2}$ and at least one of $P_1, \ldots, P_{t+1}$. By the definition of $f_t$, the coalition members can infer, using the value of the function and their input values, the input values $z_1, \ldots, z_{t+2}$. Therefore, they do not get any additional information from the execution of protocol.

This completes the proof of the theorem. □

Combining Theorem 1 with the $\left\lceil \frac{N-1}{2} \right\rceil$ - private protocols of [1, 5], we get

**Corollary 2:** The privacy hierarchy of functions defined over finite domains consists of exactly $\left\lceil \frac{N}{2} \right\rceil$ (non-empty) levels.
4 Concluding Remarks

In proving that $f_t$ is not $t + 1 -$ private, we used a partition argument (Lemma 1). We demonstrated a partition of \{1, 2, ..., $N$\} to the sets $S, \bar{S}$ with $|S| = t + 1$, such that the induced two argument function is not $1 -$ private (by Lemma 2). All known proofs of non $t -$ privacy for functions with finite domain ($t \geq \left[\frac{N}{2}\right]$) are based on a similar partition argument, together with either Lemma 2 or the general characterization of [9, 2]. It is an open problem whether such argument always suffices. That is, whether non $1 -$ privacy can always be proved by a partition argument. In particular, in [8] there are examples of (symmetric) functions which do not contradict the partition arguments, and yet are not known to be $\left[\frac{N}{2}\right] -$ private.

It is also interesting to know what is the situation with respect to functions defined over infinite domains. Clearly, the privacy hierarchy for the infinite case contains at least as many levels as the privacy hierarchy in the finite case. However, in the infinite case the hierarchy contains at least one more level: in [6] it is proved that there are functions (over countable domains) which are not even 1 - private. The existence of functions, over infinite domains, which are $t -$ private but not $t + 1 -$ private, for $1 \leq t < \left[\frac{N-1}{2}\right]$, remains an open problem.

References


