# Security and complexity aspects of Human Interactive Proofs

Nick Hopper

December 3, 2001

## 1 Introduction

My interest in Human Interactive Proofs (HIPs) is in the area of formally describing and rigorously proving the desirable properties of these systems. In particular, I am interested in the security properties of authentication-style HIPs. However I also believe that a "complexity-theoretic" definition of HIP may lead to interesting characterizations of the power of human computing.

## 2 Security and formalizations of HIPs

The HIP group at CMU has been operating under the loose definition that a HIP is a protocol that allows a human to prove something to a computer. However we have not to this point been able to produce a formal definition that is satisfactory for a number of reasons. I'll try to briefly summarize some of the difficulties and then show a candidate definition and how both the HUMANOIDS and CAPTCHA problems satisfy this definition.

### 2.1 HUMANOIDS and CAPTCHA

The HUMANOIDS project can be described briefly as a search for single-user authentication protocols which are cryptographically secure and can be executed by an unaided human. We briefly recall some definitions from [2].

**Definition 1** *An* identification protocol *is a pair of probabilistic interactive programs $(H, C)$ with shared auxiliary input $z$, such that the following conditions hold:*

- *For all auxiliary inputs $z$, $Pr[\langle H(z), C(z)\rangle = \texttt{accept}] > 0.9$*
- *For each pair $x \neq y$, $Pr[\langle H(x), C(y)\rangle = \texttt{accept}] < 0.1$*

*When $\langle H, C\rangle = \texttt{accept}$, we say that $H$* verifies his identity to *$C$, $C$* authenticates *$H$, or $H$* authenticates to *$C$.*

**Definition 2** *A protocol* $(H, C)$ *is said to be* $(\alpha, \beta, t)$ - human executable *if at least a* $(1 - \alpha)$ *portion of the human population can perform the computations* $H$ *unaided and without errors in at most* $t$ *seconds, with probability greater than* $1 - \beta$.

**Definition 3** *An identification protocol* $(H, C)$ *is* $(p, k)$-secure against active adversaries *if for all computationally bounded adversaries* $\mathcal{A}$,

$$Pr[\langle \mathcal{A}(T^k(\mathcal{A}, H(z), C(z))), C(z)\rangle = \texttt{accept}] < p \ ,$$

*where* $T^k(\mathcal{A}, H(z), C(z))$ *denotes a random variable sampled from* $k$ *sessions where* $\mathcal{A}$ *is allowed to observe and make arbitrary changes to the communications between* $H$ *and* $C$.

A brief, formal description of the HUMANOIDS project is then the search for an $(\alpha, \beta, t)$-human executable identification protocol which is $(p, k)$ secure against active adversaries for "good" values of $\alpha, \beta, t, p, k$. The focus has been on very good values of $(p, k)$ — for example, $p < 10^{-6}$ and $k > 10^6$ — and "acceptable" levels of human executability.

The CAPTCHA project which will be described by others in this workshop aims to securely authenticate users as humans (as opposed to robots.) The paradigm for these HIPs are to require the prover to show ability to solve some "AI-hard" problem.

## 2.2 Definition

We might be tempted to just extend this property to any interactive proof system $(P, V)$, and say that $(P, V)$ is an $(\alpha, \beta, t)$-HIP if $(P, V)$ is an IP and one of $P$ or $V$ is $(\alpha, \beta, t)$-human executable; and say that $(P, V)$ is a HIP if $(\alpha, \beta, t)$ pass some minimum reasonable requirement. The problem here is that both HUMANOIDS and CAPTCHA are not really interactive proofs in the same sense as the traditional IP. Such proofs are normally of the form $x \in L$ for some language $L$. But it's not really clear what $x$ or $L$ are for our projects.

What we're really talking about in these protocols is a "private key" interactive proof. Both $H$ and $C$ have some "secret" input, and $H$ wants to show $C$ that their auxiliary inputs satisfy some relation R. Note that this is different from the usual IP, which assumes an all-powerful P (no need for aux. input, he can just compute it), and ZK, which assumes that P may have an auxiliary input but not V. So in essence we want to prove that we're "talking about the same thing." The following complexity-style definitions for HIPs are really just straightforward translations of this idea.

**Definition 4** *A pair of interactive programs* $(P, V)$ *with auxiliary inputs is an* Interactive Proof of Related Secrets *for the relation* $R$ *if:*

- *For all* $(x, y) \in R$, $Pr[\langle P(x), V(y)\rangle = \texttt{accept}] > 1 - \epsilon$.
- *For all* $(x, y) \notin R$, $Pr[\langle P(x), V(y)\rangle = \texttt{accept}] < \epsilon$.

- *For any $\mathcal{P}$, let $p = Pr[\langle \mathcal{P}, V(y) \rangle = \texttt{accept}]$. Then if $p > \epsilon$, there exists an $x$ such that $(x, y) \in R$ and $\mathcal{P}$ and $P(x)$ are computationally indistinguishable.*

So having a "cheating" prover for some $y$ is as good as having a legitimate prover. A technical issue is that we may wish to consider altering this definition to say that there's a program $E$ which given oracle access to a successful cheating prover $\mathcal{P}$ is computationally indistinguishable from some $P(x)$, where we require that the running time of $E$ is $O(1/|p - \epsilon|^c)$.

We then say that an Interactive Proof of Related Secrets is a Human Interactive Proof if the prover is $(\alpha, \beta, t)$-human executable for some reasonable threshold of $(\alpha, \beta, t)$. Of course this is not a rigorous definition unless we have a rigorous model of human computation, but it can definitely be established empirically. Both the HUMANOIDS and CAPTCHA protocols are meant to additionally satisfy the requirement of $(p, k)$-security against active adversaries given in the previous section; we say that such HIPs are Secure HIPs. Without the security requirement a HIP can be thought of as an automatically gradable, "uncheatable" test.

Given this definition, it's not too hard to fit CAPTCHA and HUMANOIDS in as HIPs for specific relations and having additional properties, analogous to the "Zero-Knowledge" property of some interactive proofs.

### 2.2.1 HumanOIDS

Under this definition, a HUMANOIDS simply becomes a HIP for the "equality" relation, with the additional security requirements that no computationally bounded adversary can authenticate with probability better than $p$ even after seeing $k$ interactions - this is the condition of $(p, k)$-*security against passive adversaries* given in the Asiacrypt paper.

### 2.2.2 CAPTCHA

A CAPTCHA is then a HIP for the relation "$x = y^{-1}$" for programs $x$ and $y$, with the additional requirements that most humans can invert $y$ and writing a program to invert $y$ is an AI-hard problem. The security condition in the definition of a HIP ensures us that there's no way to "trick" a verifier into accepting without really having solved the AI problem.

## 3 HIP Complexity

In complexity theory, the study of interactive proofs has led to many novel characterizations for the computational properties of several complexity classes; e.g., PSPACE in terms of IP, NP in terms of PCP [1]. Assuming we accept the definitions above for HIP, or similar definitions, then the study of what relations have HIPs would seem to have potential for giving us a novel characterization of the computational abilities of unaided humans.

Thus I am interested in several of the following questions:

- The general question of the expressive power of HIPs. What relations have HIPs? What other specializations of HIPs are useful?
- Towards impossibility - Human Executability is not exactly transitive, but it's conceivable that we could produce a notion of reducibility between HIPs. Is there a HIP-complete relation?
- A consequence of the rather weak security condition is that any relation which is human-computable has a HIP with human verifier: on input $x$, $P$ says $x$ and $V$ checks $(x, y) \in R$. So perhaps this is just shifting the question to: "what relations are human computable?" Certainly $=$ and $>$ are two. For small sets, $\subseteq$ is also human computable; this might be useful for group membership.

# References


[1] Goldreich, Oded. Modern Cryptography, Probabilistic Proofs and Pseudorandomness. Algorithms and Combinatorics Series (Vol. 17), Springer, 1998.

[2] Hopper, Nicholas J. and Manuel Blum. Secure Human Identification Protocols. In Colin Boyd, ed.: Advances in Cryptology - ASIACRYPT 2001. LNCS Volume 2248 (2001) pp. 52 – 67.