CS276 Midterm 1 Common Errors

March 28, 2004

Omitting a reduction. The most common error was to be too informal in the proof. When proving that the security of F implies the security of G, you almost always need to give an explicit reduction: show how any given attack on G can be used to build an attack on F. It's the reduction that protects you from making subtle errors of reasoning.

The correlation fallacy. A common error was to overlook the possibility of correlation. Suppose I have three random variables X, Y, Z, and I know that $X \sim Y$ (i.e., X, Y are indistinguishable). What can I conclude? Here is the fallacy:

> **Claim (bogus).** If $X \sim Y$, then $(X, Z) \sim (Y, Z)$.
>
> **"Proof" (bogus).** The only difference between (X, Z) and (Y, Z) is in the first half of the pair. Consequently, any attacker can only distinguish between these two by distinguishing between X and Y. But we assumed that X and Y are indistinguishable, hence the claim follows. $\square$

The above reasoning is intuitively appealing, but fatally flawed. Maybe the distinguisher can somehow exploit the correlation between the first and second half of the pair; the above reasoning fails to take that possibility into account.

We can easily come up with a counterexample. Let $G : \{0,1\}^n \to \{0,1\}^\ell$ be a secure pseudorandom generator, and choose $k \leftarrow \{0,1\}^n$ and $u \leftarrow \{0,1\}^\ell$ uniformly and independently at random. Let $Z = k$, $X = G(k)$, and $Y = u$. It is easy to see that $X \sim Y$, since $G(k) \sim u$. At the same time, $(X, Z) \not\sim (Y, Z)$, since $(G(k), k)$ is easily distinguished from (u, k). The distinguisher: on input (a, b), simply check whether $G(b) \stackrel{?}{=} a$.

One common place where this fallacy manifested was in Problem 1, where some people claimed that, since G is a (t, ϵ_G)-secure PRG, $(E(k), G(k))$ is ϵ_G-indistinguishable from $(E(k), u)$. Their reasoning invoked the fallacy above, noting that the first half of the pair is the same in both cases, and so the only difference is in the second half of the pair. In fact, the claim is wrong, as may easily be seen by considering the identity encryption scheme $E(k) = k$; then $(E(k), G(k))$ is trivially distinguishable from $(E(k), u)$, no matter how secure G is.

Another place where this fallacy manifested was in hybrid arguments, where people argued that the only difference between the i-th and $i+1$-th hybrid was in the response to the i-th oracle query.

The reason people fell for this is that they omitted the reduction. If you try to write down an explicit reduction, you quickly discover the flaw in the reasoning.

For reference, here is the corrected version of the above.

> **Lemma (corrected).** Let X, Y, Z be random variables, where Z is independent of both X and Y, and where there is an efficient algorithm to sample from the distribution

of Z in time t'. If X and Y are (t, ϵ)-indistinguishable, then (X, Z) and (Y, Z) are $(t - t' - O(1), \epsilon)$-indistinguishable.

Proof. Consider any algorithm B that runs in time $t - t' - O(1)$ and tries to distinguish between (X, Z) and (Y, Z). Define the algorithm A by

$A(s)$:
1. Choose t randomly according to the distribution on Z.
2. Run $B(s, t)$, and return whatever B does.

Notice that $A(X)$ has the same distribution as $B(X, Z)$, and $A(Y)$ has the same distribution as $B(Y, Z)$. Also, A runs in time at most t. Then we find

$$\Pr[B(X, Z) = 1] - \Pr[B(Y, Z) = 1] = \Pr[A(X) = 1] - \Pr[A(Y) = 1] \leq \epsilon,$$

or in other words, A's distinguishing advantage is at most ϵ. $\square$

Distinguishing is not an event. A few people wrote something like "Let E be the event that A distinguishes $G(k)$ from random," and then proceeded to try to condition on $\neg E$ or somesuch. The problem is that the quoted statement doesn't make any sense; breaking G is not an event, and we can't apply bad-event style analysis to something that's not an event.

Recall the definition of an event. In basic probability theory, our probability space is defined by a set Ω of outcomes and a probability function $\Pr : \Omega \to [0, 1]$ satisfying certain properties. An event E is a set of outcomes, namely, a subset of Ω. In the above example, the sample space is the set of values for k (say); but there is no set of such values for which we can identify A as successfully distinguishing $G(k)$ from random. So this idea doesn't work, alas.

Trying to prove too much. A few people tried to prove, in Problem 1, that if either E is RR-secure or G is a secure PRG, then $\mathcal{E}$ will be RR-secure. But there is no hope of proving this, as is easily seen by considering the identity encryption scheme $E_{\text{pk}}(m) = m$ along with some secure PRG, say $G =$ AES-CTR; then we find that $\mathcal{E}$ is totally insecure, even though G is secure.

Union vs. hybrid arguments. Also, I noticed a few people confusing the union bound and hybrid arguments. If we want to show, for instance, that $(G(k_1), \ldots, G(k_q)$ is $q\epsilon$-pseudorandom if G is ϵ-pseudorandom, we're using a hybrid argument, not a union bound.

RR implies LR. Another (minor) error was to note that we proved in Lecture 6 that RR $\Rightarrow$ LR, but to get the parameters wrong. The concrete security parameters don't appear explicitly in the lecture notes, but it is pretty easy to figure out that (t, q, ϵ)-RR-security implies $(t - O(q), q, 2\epsilon)$-LR-security. Some people forgot the factor of two. Not a big deal.

Failure of imagination. The worst error was to try to "prove" security of a scheme S by listing a few possible attacks and then blindly asserting that these are the only conceivable way to break S, so it must suffice to prove that these attacks don't work. Such an argument is unconvincing. For all we know, maybe there is some other way to attack S, and our inability to conceive of that attack represents a failure of imagination on our part. The need to avoid this kind of ill-founded reasoning is exactly the reason why we turn to proofs, reductions, complexity theory, and all the other material covered in this class.