Obfuscation

Lecture 25

Obfuscation

- The art & science of making programs "unintelligible"

```
#define _ -F<00||--F-OO--;
int F=00,OO=00;main(){F_OO();printf("%1.3f\n",4.*-F/OO/OO);}F_OO()
{
            _-_-_-_
       _-_-_-_-_-_-_-_-_
    _-_-_-_-_-_-_-_-_-_-_-_
  _-_-_-_-_-_-_-_-_-_-_-_-_-_
 _-_-_-_-_-_-_-_-_-_-_-_-_-_-_
 _-_-_-_-_-_-_-_-_-_-_-_-_-_-_
_-_-_-_-_-_-_-_-_-_-_-_-_-_-_-_
_-_-_-_-_-_-_-_-_-_-_-_-_-_-_-_
_-_-_-_-_-_-_-_-_-_-_-_-_-_-_-_
_-_-_-_-_-_-_-_-_-_-_-_-_-_-_-_
 _-_-_-_-_-_-_-_-_-_-_-_-_-_-_
 _-_-_-_-_-_-_-_-_-_-_-_-_-_-_
  _-_-_-_-_-_-_-_-_-_-_-_-_-_
    _-_-_-_-_-_-_-_-_-_-_-_
        _-_-_-_-_-_-_-_
            _-_-_-_
}
```

from International Obfuscated C Code Contest 1988 (via Wikipedia)

  - The program should be fully functional
  - It may contain secrets that shouldn't be revealed to the users (e.g., signature keys) — any more than executing it reveals

Obfuscation

- For protecting proprietary algorithms, for crippling functionality (until license bought), for hiding potential bugs, for hardwiring cryptographic keys into apps, for reducing the need for interaction with a trusted server (say for auditing purposes), ...
- Several heuristic approaches to obfuscation exist
  - All break down against serious program analysis

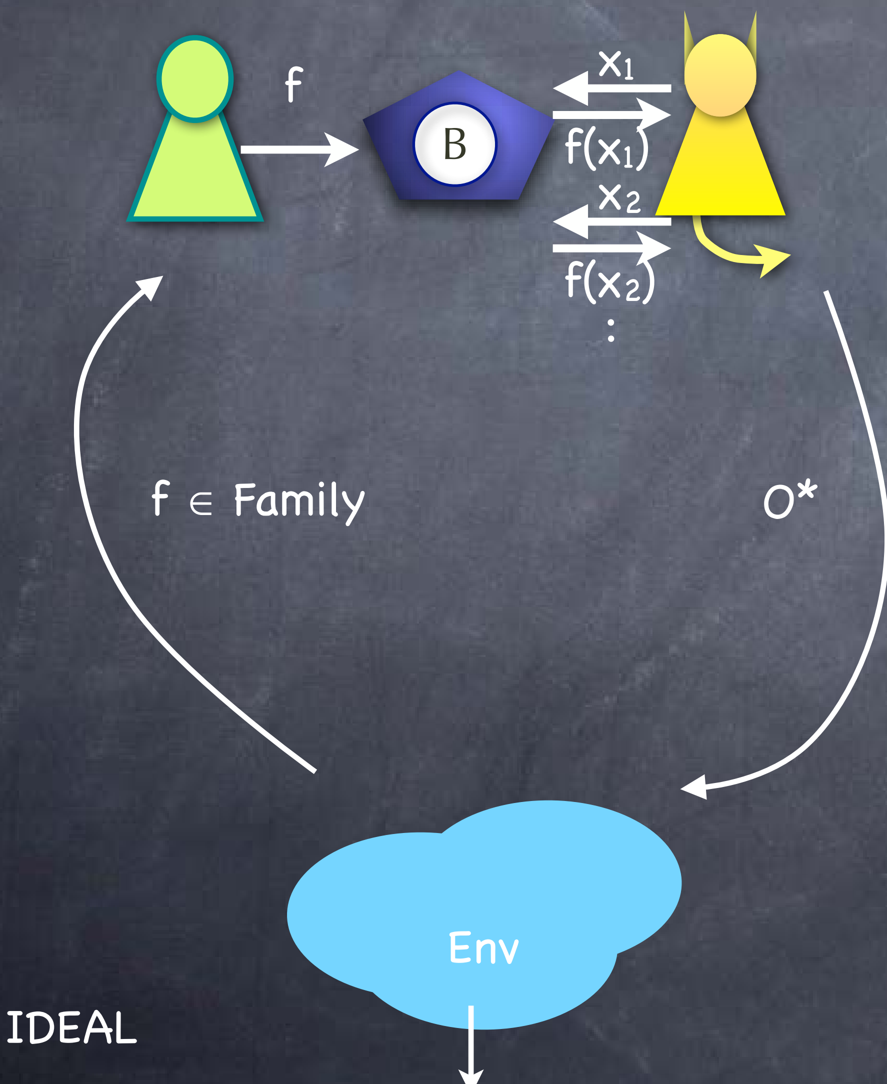
Cryptographic Obfuscation

- Obfuscation using cryptography?
  - Need to define a security notion
  - Constructions which meet the definition under computational hardness assumptions
- Cryptography using obfuscation
  - If realized, obfuscation can be used to instantiate various other powerful cryptographic primitives
  - Example: PKE from SKE. Obfuscate the SKE encryption program with the key hardwired (plus a PRF for generating randomness from the plaintext), and release as public-key
    - Or FE: Encrypt message x with a CCA-secure PKE. Function key SK_f is a program that decrypts, computes f(x) and outputs it.

Defining Obfuscation: First Try

Note: Considers only corrupt receiver

Too strong! Requires family to be <u>learnable</u> from black-box access



Secure (and correct) if:

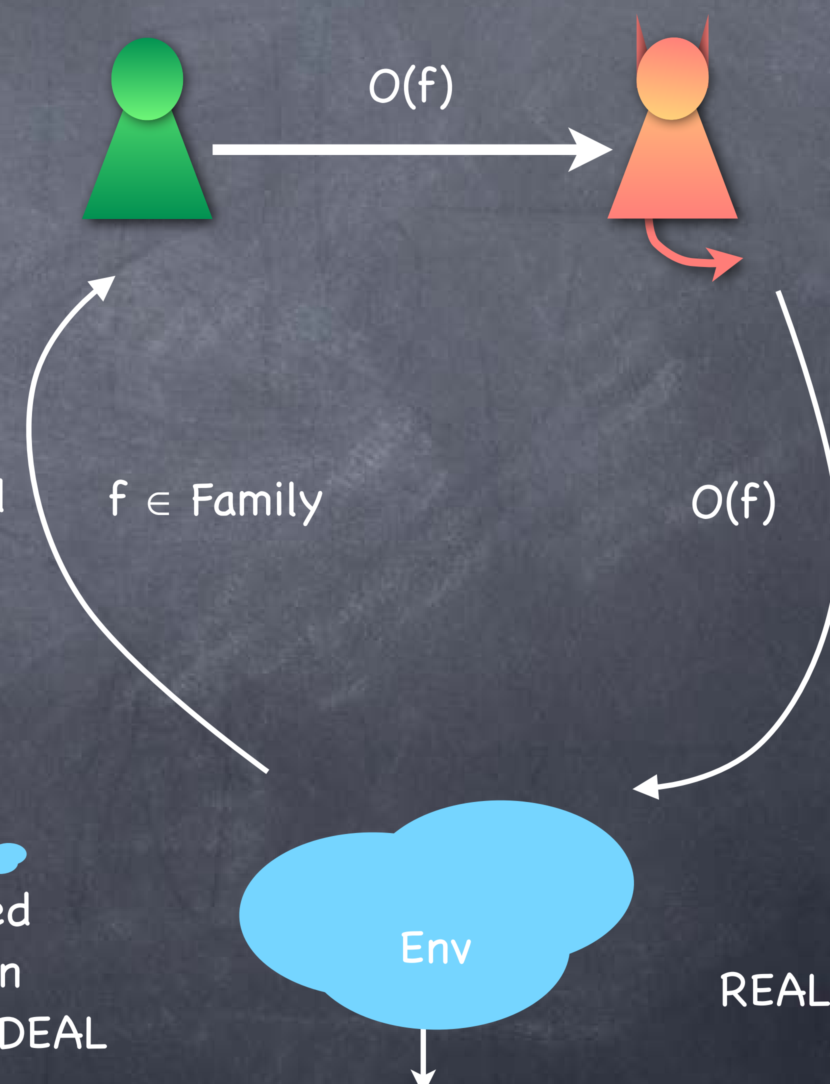
∀

∃ s.t.

∀

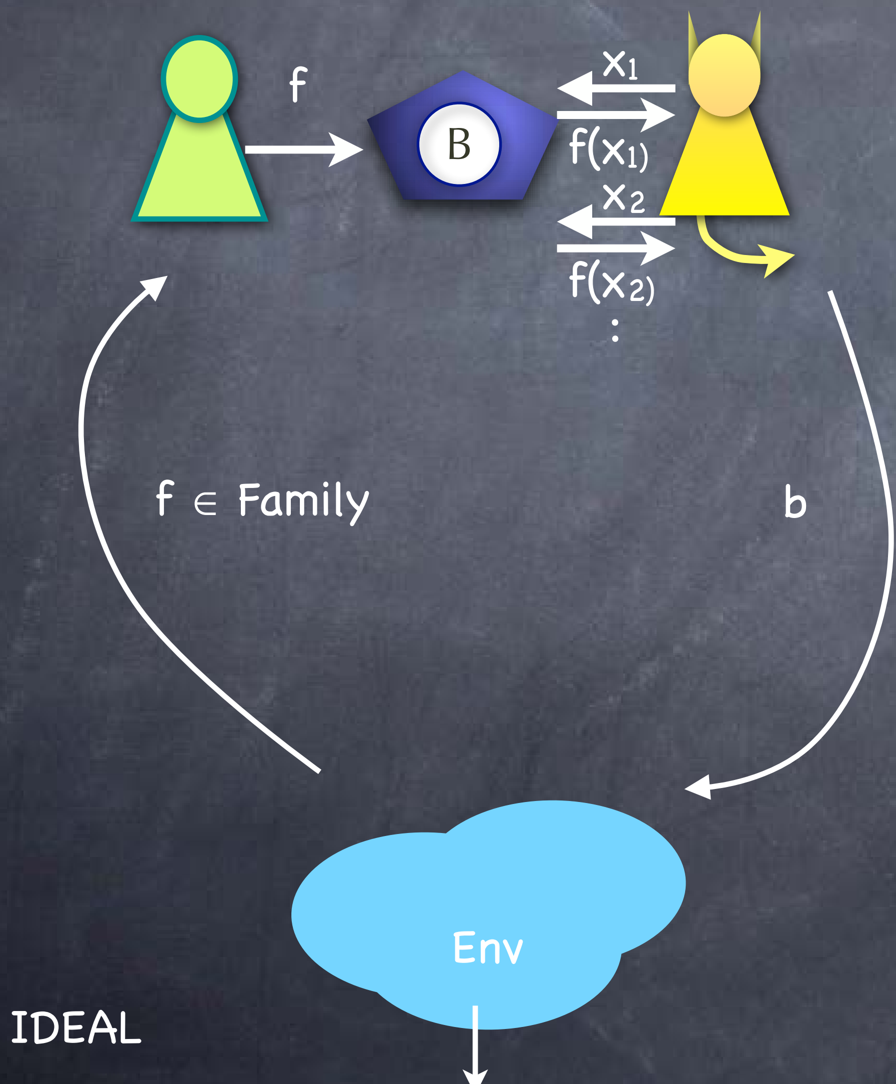
output of

is distributed identically in REAL and IDEAL



Defining Obfuscation: First Try

Note: Considers only corrupt receiver



Virtual Black-Box (VBB) Obfuscation

Secure (and correct) if:

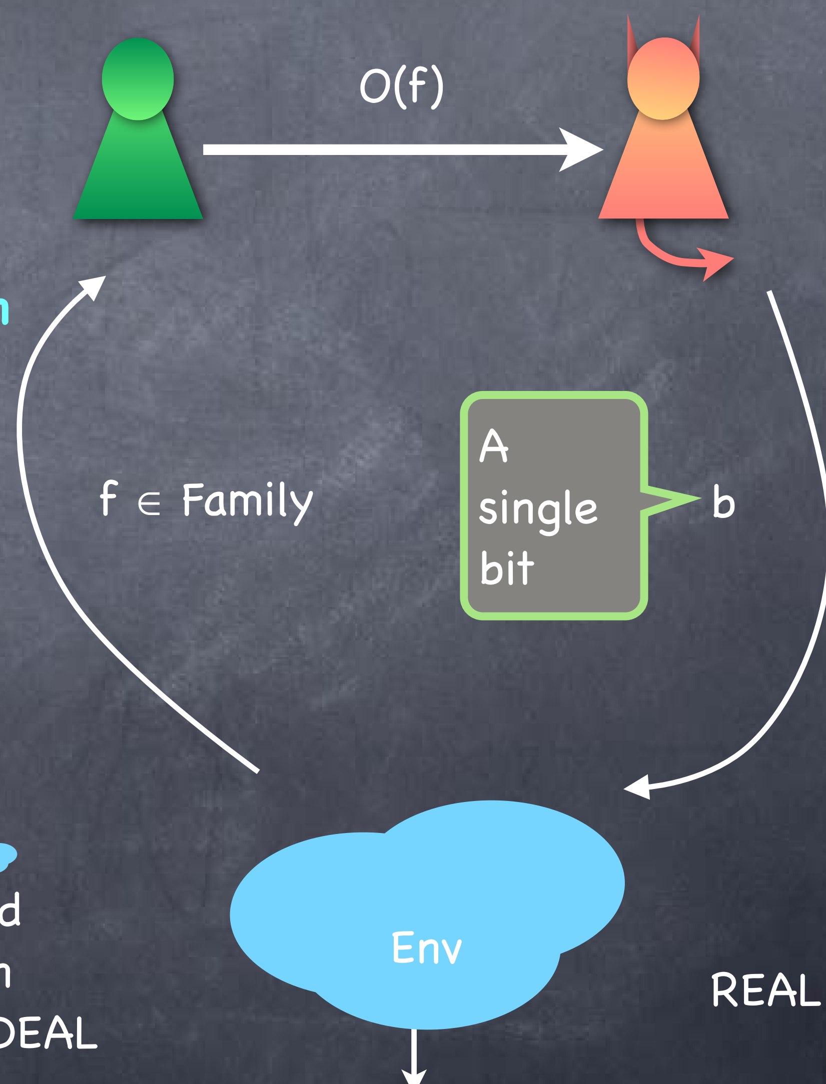
∀

∃ s.t.

∀

output of

is distributed identically in REAL and IDEAL



Impossibility of Obfuscation

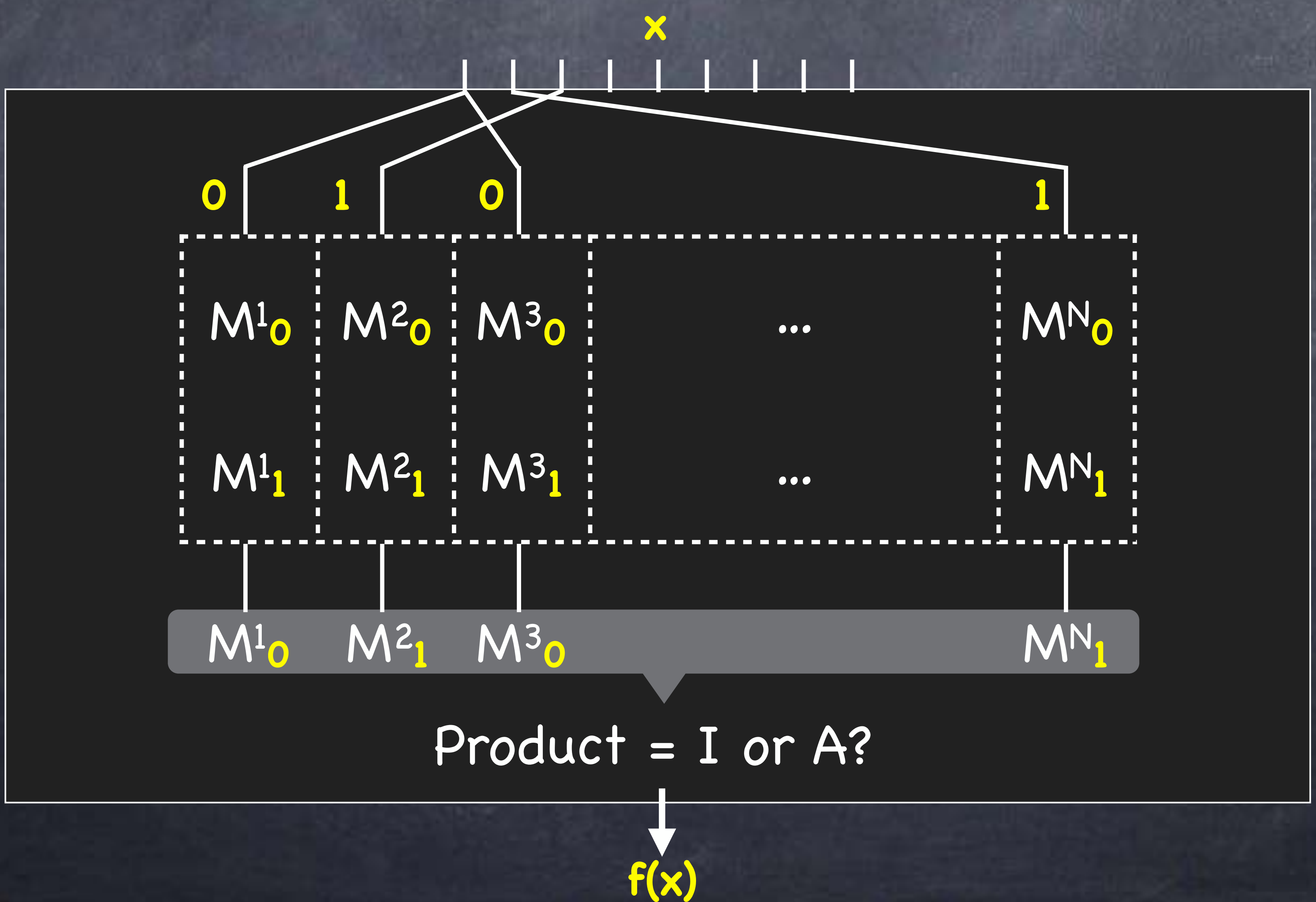
- VBB obfuscation is impossible in general
- Explicit example of an unobfuscatable function family
    - Idea: program which when fed its own code (even obfuscated) as input, outputs secrets
    - Programs $P_{\alpha,\beta}$ with secret strings α and β:
        - If input is of the form $(0,\alpha)$ output β
        - If input is of the form $(1,P)$ for a program P, run P with input $(0,\alpha)$ and if it outputs β, output (α,β)
    - When $P_{\alpha,\beta}$ is run on its own (obfuscated) code, it outputs (α,β). Can learn, e.g., first bit of α. In the ideal world, need to guess!

Possibility of Obfuscation

- Hardware assisted
- For simple function families
  - e.g., Point functions (from perfectly one-way permutations)
  - But general "low complexity classes" are still unobfuscatable (under cryptographic assumptions)
- For weaker definitions
- In idealized models (random oracle model, generic group model, etc.)
  - Need a suitable representation of the function

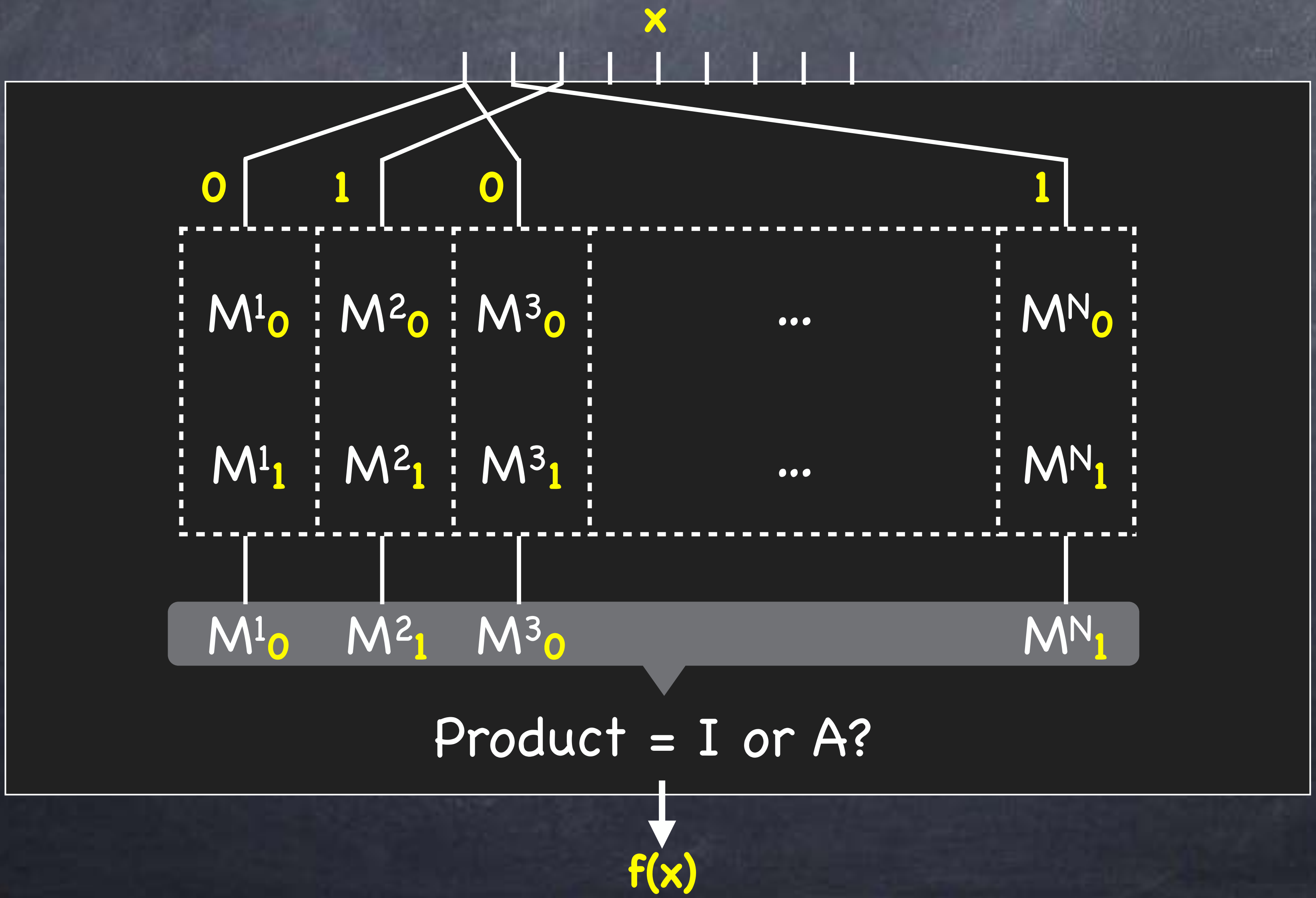
Matrix Programs

- $f : \{0,1\}^n \rightarrow \{0,1\}$ using a set of 2N w×w matrices (N = poly(n))
- Family F: all f in F have the same N, w, matrix A and "wiring"



Matrix Programs

- To obfuscate, encode matrices s.t. only valid matrix multiplications and final check can be carried out (for any x)
- No other information about the 2N matrices should be deducible



Multi-Linear Map

- Recall groups with bilinear pairing:
  - e: $G_1 \times G_2 \to G_T$ such that $e(g_1^a,g_2^b) = g_T^{ab}$
  - Also group operations in G_i
  - I.e., one multiplication and several additions (in the exponent)
  - Assumption: Hard to carry out other operations like $(g_1^a,g_1^b) \mapsto g_T^{ab}$. Heuristic: the Generic Group Model
- Extension to more than 2 groups?
  - Let $T = \{1,\ldots,k\}$. For each non-empty subset $S \subseteq T$, a group G_S.
  - $e(g_{S_1}^a,g_{S_2}^b) = g_{S_3}^{ab}$, where $S_1 \cap S_2 = \emptyset$ and $S_3 = S_1 \cup S_2$

Multi-Linear Map

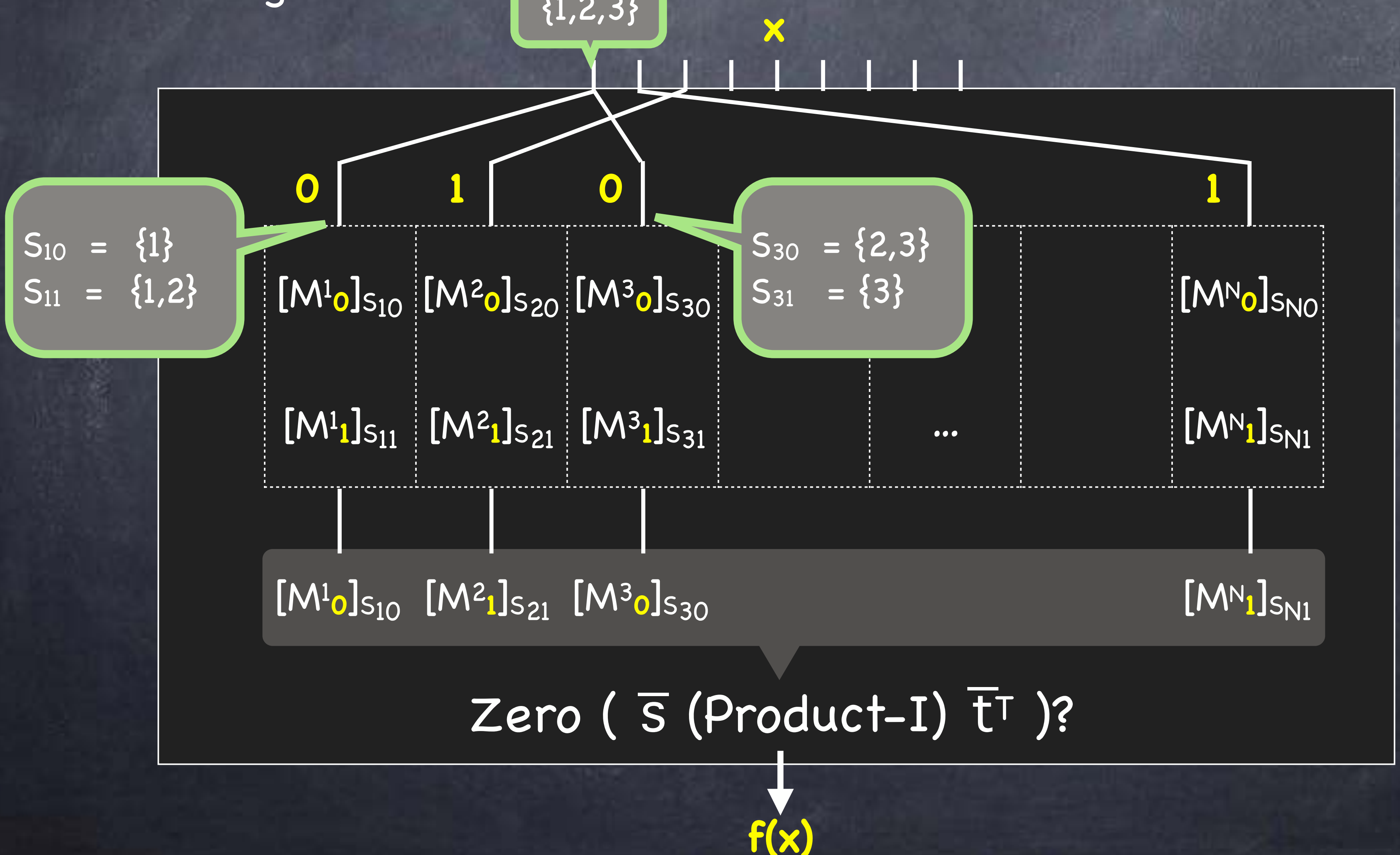
- Let $T = \{1,\ldots,k\}$. For each non-empty subset $S \subseteq T$, a group G_S.
- An element a encoded in G_S: $[a]_S$ (think g_S^a)
  - Need a private key for encoding (think of keeping g_S secret)
  - Allowed to learn the set S from $[a]_S$
- Following public operations:
  - $[a]_S + [b]_S \rightarrow [a+b]_S$ (note that S is the same for all)
  - $[a]_{S_1} * [b]_{S_2} \rightarrow [ab]_{S_1 \cup S_2}$ where $S_1 \cap S_2 = \emptyset$ and $S_3 = S_1 \cup S_2$
  - Zero-Test($[a]_T$) checks if a=0 or not (note: only for set T)
- Generic Group Model heuristic: No other operation possible!

Obfuscation from Multi-Linear Map

- Matrix elements are encoded using the multi-linear map, so that matrix product can be carried out on encoded elements
  - Final outcome checked as $[a]_T = [v]_T$, where $[a]_T$ is computed and $[v]_T$ is included as part of the obfuscation
- Each matrix encoded using an associated set S
  - Sets chosen so as to prevent invalid combinations
  - Matrices randomized (while preserving product) to ensure that the matrices cannot be reordered/tampered with
    - Any tampering will result (w.h.p.) in $[a]_T$ being random (and independent each time)

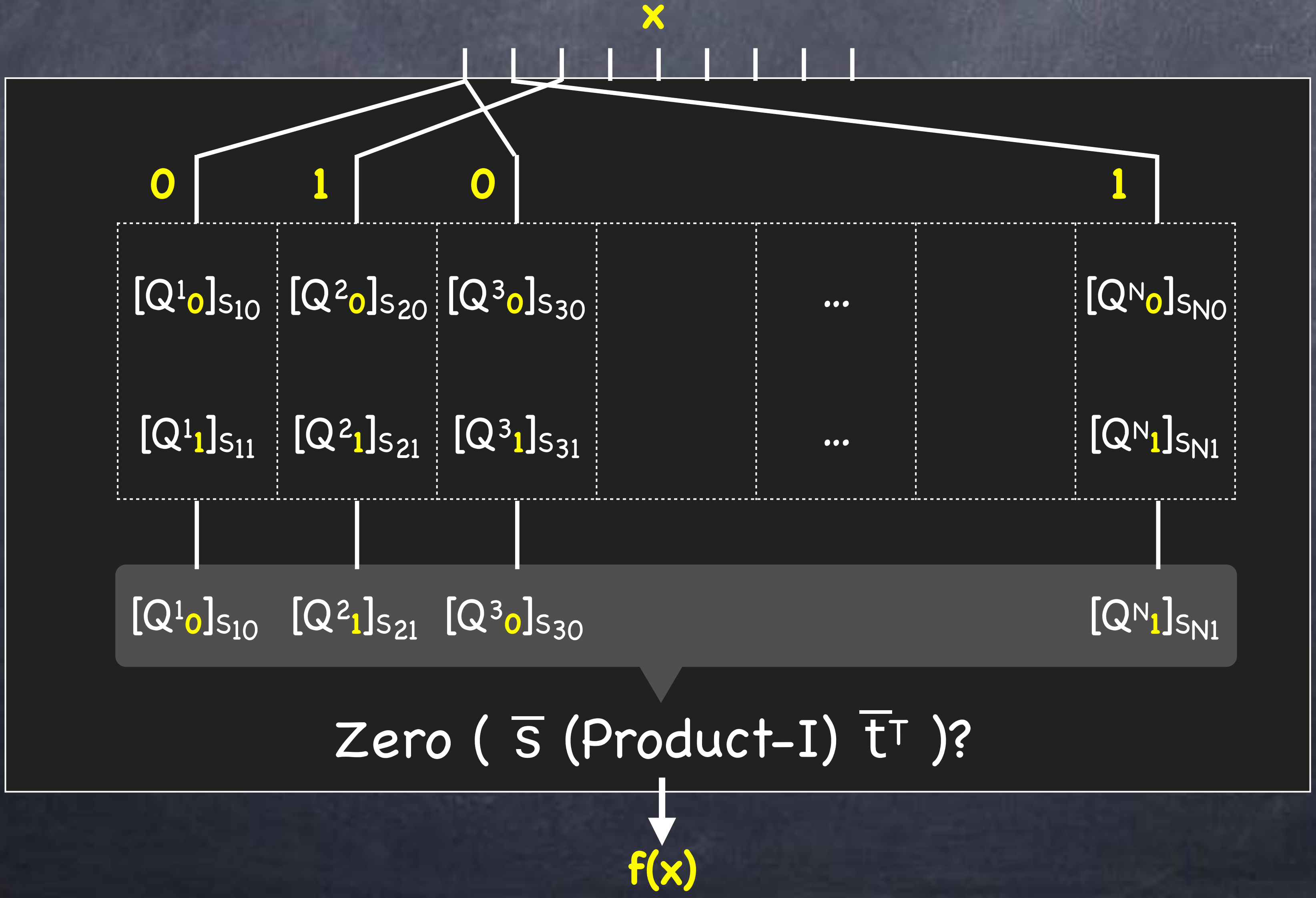
Obfuscating Matrix Programs

- Preventing invalid combinations: entries in $M^i_{0/1}$ encoded for set $S^i_{0/1}$ so that invalid combinations result in intersecting sets, or sets not covering T



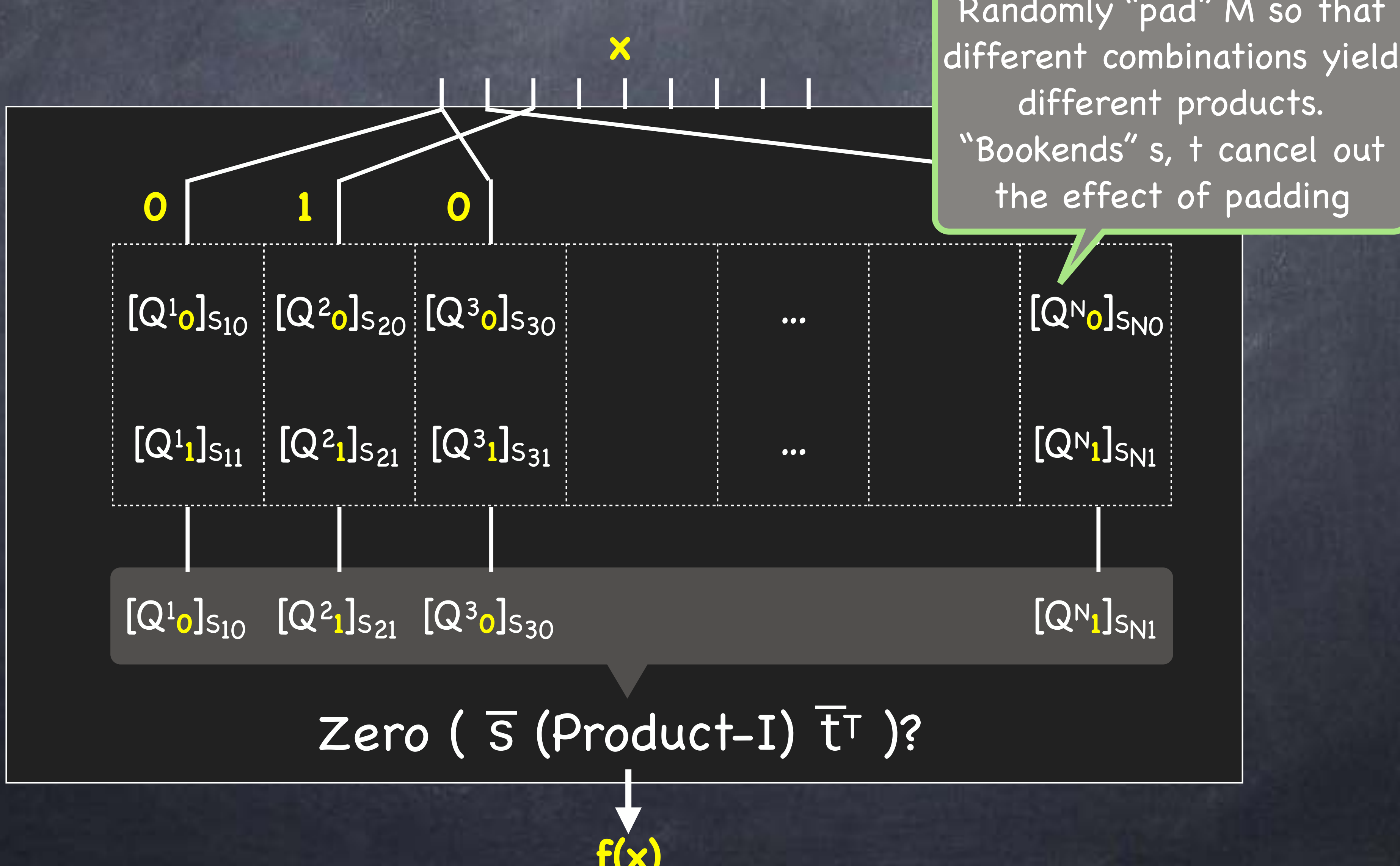
Obfuscating Matrix Programs

- Ensure no information by reordering/tampering with the matrices
  - Let $Q^i_b = R_{i-1} M^i_b R_i^{-1}$ (R_i random, $R_0=R_N=I$): $\prod_i Q^i_{b_i} = \prod_i M^i_{b_i}$ while $\{Q^i_{b_i}\}$ has no information about $\{M^i_{b_i}\}$ than its product



Obfuscating Matrix Programs

- Partial evaluations: If two inputs yield same matrix product for a "segment" of the computation, the encodings obtained will also be same



Obfuscating Matrix Programs

- Using generic multi-linear map, this yields Virtual Black-Box obfuscation for polynomial-sized matrix programs
- Barrington's Theorem: "Shallow" circuits (NC^1 functions) have polynomial-sized matrix programs (with 5x5 matrices)
    - Can "bootstrap" from this to all polynomial-sized circuits/ polynomial-time computable functions, assuming Fully Homomorphic Encryption (with decryption in NC^1)
- Do multi-linear maps exist?
    - Generic multi-linear map model is an unrealizable model (because VBB obfuscation for NC^1 is impossible!)
    - Weaker multi-linear maps?

Obfuscating Matrix Programs

- Several candidate multi-linear maps proposed [GGH'13, CLT'13,...]
  - Have noisy, randomized encoding
  - Initial candidates already broken...
- Instantiating obfuscation constructions using these candidates yield weaker forms of obfuscation (in standard model)
- Indistinguishability Obfuscation (iO), Differing-Inputs Obfuscation (DIO), etc.
  - Weaker, but still useful in many applications
  - Underlying security notion: "Indistinguishability-Preserving"
- Recent alternate constructions of iO via "Compact FE" can be based on L-linear maps with L possibly as low as 3

Today

- Obfuscation
- Strong definitions are provably impossible to achieve
- Recent breakthroughs (for weaker definitions)
  - Using Multi-linear Maps
  - Still being cryptanalyzed