Protocol Skeletons and Statistics

Statistical Analysis

- Revisit "Information Theoretic" model of information security
- Consider run of protocol(s) as a stochastic process

$$\mathbf{X}_k \quad k \in \text{nodes}$$

- Random variable $\mathbf{X}_k$ is bitstring valued

$$\mathbf{X}_k : \Omega \to \mathcal{Y}_k = \{0, 1\}^{\ell(k)}.$$

Fashion Statement

- Information Theoretic Protocol research:
  - Démodé,
  - Désuet.
- In many known instances, protocols are informational theoretically vulnerable,

Outline and Purpose of Talk

- Relationship between protocol analysis and classical notions of conditional expectation and information.
- Recent work on skeletons suggest relationship between strand spaces and processes
- Review basic terminology.
    - Probability and Statistics
    - Skeletons.

Stochastic Process

- Family of random variables

$$\mathbf{X}_k : \Omega \to \mathcal{E} \quad k \in I = \text{index set}$$

- The value space $\mathcal{E}$ is often $\mathbb{R}$.
  - Any complete separable metric space is acceptable.
- Index set is (discrete or continuous) time or space.
- A "statistic" is some function of $\mathbf{X}_k$.
- Examples:
  - Time average:

$$\text{Avg}_N = \frac{\left(\mathbf{X}_1 + \cdots + \mathbf{X}_N\right)}{N}$$

  - Maximum

$$\text{Max}_N = \max\left(\mathbf{X}_1, \ldots, \mathbf{X}_N\right)$$

Statistic

- A "statistic" on node k is a function on $\mathcal{Y}_k$.
  - Linear statistic $\mathcal{Y}_k \to \mathbb{Z}/(2)$:

$$\mathsf{L}_F(x) = \text{parity of card}(F \cap x)$$

where

$$F \subseteq \{1, 2, \ldots, \ell(k)\} \text{ a bitpattern}$$

L_F is linear as a vector space map $\mathcal{Y}_k \to \mathbb{Z}/(2)$.

- Estimate linear statistics on node n from linear statistics on nodes $k_1, \ldots k_r$.
- Easier to consider

$$\chi_F(x) = (-1)^{\mathsf{L}_F(x)} \in \{-1, 1\}.$$

Skeletons

CryptoAlgebra

- Structure B
  - Binary (possibly partial) operators: *concatenation* and *encryption*
  - domain(B) is partitioned into
    - *atoms*: B_{atom}
    - *concatenations*, and
    - *encryptions*
- Concrete cryptoalgebras: Usually a space of bit-strings
  - Concatenation may be partial

Freeness

- Free cryptoalgebras,
  - `Put in some abstract nonsense here`
- A_{B} is the free algebra generated from $\mathsf{B}_{\mathsf{atom}}$
- There is a partial surjective map $\pi : \mathsf{A}_{\mathsf{B}} \to \mathsf{B}$
- $\mathsf{A}_{\mathsf{B}}[\overline{X}]$ is the free algebra generated from $\mathsf{B}_{\mathsf{atom}}$ and the set of parameters $\overline{X}$.
- Machinery of authentication tests assumes message algebra is free.

Parametric Skeletons

- prot is a protocol
- $\overline{X}$ is a set of parameters,
- *unordered skeleton* for prot, $\overline{X}$
  - A set of $\mathcal{R}$ regular strands of prot,$\overline{X}$, and
  - A "stopping index" $h\colon \mathcal{R} \to \mathbb{N}$ is such that $s \in \mathcal{R}$ implies $h(s) \leq \text{length}(s)$.
- *ordered skeleton* $\mathbb{A}[\overline{X}]$:
  - unordered skeleton with a weak partial order on nodes compatible with the strand order.

Parameters and Genericity

- Some of the parameters in $\overline{X}$ are cryptoparameters: Nonces and keys.
- A set of regular strands S is generic for a protocol: For any cryptoparameter X
  - X can originate only once in S,
  - If X is a secret parameter, then X originates nowhere in S.

Random Skeletons

- Generic parametric skeleton $\mathbb{A}[\overline{X}]$.
- Space of values $\mathfrak{X}$ of parameters has probability measure $\mathsf{P}_{\mathfrak{X}}$.
- For each node $n \in \mathbb{A}[\overline{X}]$, $\mathsf{term}(n)$ is message template with parameters in $\overline{X}$.
- Plugging in values for parameters in $\mathsf{term}(n)$ determines a function $\mathbf{X}_n$ on $\mathfrak{X}$.
- $\mathbf{X}_n$ is a random function with values in bit string space $\mathcal{Y}_n$.

Partitions and Algebras of Sets

- $\mathbf{Y} : \mathfrak{M} \to \mathcal{Y}$ is a (finite range) random variable.

$$\mathcal{A}_{\mathbf{Y}} = \{\mathbf{Y}^{-1}(B) : B \subseteq \mathcal{Y}\}$$

- $\{\mathbf{Y}^{-1}(c) : c \in \mathcal{Y}\}$ partitions $\mathfrak{M}$.
- Corresponds to the distinctions in $\mathfrak{M}$ that can be made by observing $\mathbf{Y}$.
- Easily generalizes to metric space valued random variables on any probability space.

Classical Dependency

- $\mathbf{Y}$, $\mathbf{Z}$ are random variables with values in $\mathcal{Y}$, $\mathcal{Z}$ respectively.
- $\mathbf{Z}$ completely dependent on $\mathbf{Y}$ iff there exists a (Borel) function $g : \mathcal{Y} \to \mathcal{Z}$ such that

$$g \circ \mathbf{Y} = \mathbf{Z} \quad \text{almost surely}$$

- Theorem: $\mathbf{Z}$ completely dependent on $\mathbf{Y}$ iff

$$\mathcal{A}_{\mathbf{Z}} \subseteq_{\text{mod nullsets}} \mathcal{A}_{\mathbf{Y}}.$$

- g may be very hard to compute.
- For finite range random variables, result is completely elementary.

Distribution

- $\mathfrak{M}$ is a probability space with probability $\mathsf{P}_{\mathfrak{M}}$
- $\mathbf{W} : \mathfrak{M} \to \mathcal{W}$ a random variable.
- *Distribution of* $\mathbf{W}$, $\mathsf{P}_{\mathbf{W}}$:

$$\mathsf{P}_{\mathbf{W}}(A) = \mathsf{P}(\mathbf{W}^{-1}(A)) \quad A \subseteq \mathcal{W} \text{ Borel}$$

- W is a measure preserving transformation:

$$\mathbf{W} : \mathfrak{M} \to \mathcal{W}_{\mathsf{P}_{\mathbf{W}}}$$

- Regard random variable as a measure preserving mapping between probability spaces.

Numerical Random Variables

- $(\mathfrak{M}, \mathsf{P})$ is a probability space.
- $L^2_{\mathsf{P}}(\mathfrak{M})$ space of random variables ψ such that

$$\mathsf{E}_{\mathsf{P}}(|\psi|^2) < \infty$$

- Space is suitable for studying covariance properties of numerical random variables.
- Inner product

$$\langle \psi, \phi \rangle = \mathsf{E}_{\mathsf{P}}(\psi\phi)$$

- Length

$$\|\psi\|_{L^2_{\mathsf{P}}(\mathfrak{M})} = \sqrt{\mathsf{E}_{\mathsf{P}}(|\psi|^2)}$$

Koopmanism

- $L^2_{\mathsf{P}}(\mathfrak{M})$ a space of complex valued functions on $\mathfrak{M}$. In finite case, all functions.
- Inner product carries "covariance structure".
- For $\mathbf{W} : \mathfrak{M} \to \mathcal{W}$ a random variable.

$$\mathsf{U}_{\mathbf{W}} : L^2_{\mathsf{P}_{\mathbf{W}}}(\mathcal{W}) \to L^2_{\mathsf{P}}(\mathfrak{M})$$

  is

$$\mathsf{U}_{\mathbf{W}}(\psi) = \psi \circ \mathbf{W}.$$

- Koopmanism: $\mathsf{U}_{\mathbf{W}}$ is a linear isometry.

$$\| \mathsf{U}_{\mathbf{W}}\, \psi \|_{L^2_{\mathsf{P}}(\mathfrak{M})} = \|\psi\|_{L^2_{\mathsf{P}}(\mathfrak{M})}$$

- Compositionality: $\mathbf{W} : \mathfrak{M} \to \mathcal{W}$, $\mathbf{Z} : \mathcal{W} \to \mathcal{Z}$,

$$\mathsf{U}_{\mathbf{Z} \circ \mathbf{W}} = \mathsf{U}_{\mathbf{W}}\, \mathsf{U}_{\mathbf{Z}}$$

Conditioning

- $\mathbf{Y} : \mathfrak{M} \to \mathcal{Y}$, $\mathbf{Z} : \mathfrak{M} \to \mathcal{Z}$ are random variables.
- Conditional expectation extracts information from $\mathbf{Z}$ observed from random variable $\mathbf{Y}$.
- Conditioning is a linear map:

$$\mathsf{E}_{\mathbf{Z}|\mathbf{Y}} : L^2_{\mathsf{P}_\mathbf{Z}}(\mathcal{Z}) \to L^2_{\mathsf{P}_\mathbf{Y}}(\mathcal{Y})$$

- Theorem. Conditioning operator can be determined by Koopman operators

$$\mathsf{E}_{\mathbf{Z}|\mathbf{Y}} = \mathsf{U}^*_\mathbf{Y} \circ \mathsf{U}_\mathbf{Z}$$

Conditioning

- Random variable picture:

$$\begin{tikzcd} \mathcal{Z} & \mathfrak{M} \arrow[l, "\mathbf{Z}"'] \arrow[r, "\mathbf{Y}"] & \mathcal{Y} \end{tikzcd}$$

- Koopman picture:

$$\begin{tikzcd} L^2_{\mathsf{P}_{\mathbf{Z}}}(\mathcal{Z}) \arrow[r, "U_{\mathbf{Z}}"'] & L^2_{\mathsf{P}}(\mathfrak{M}) & L^2_{\mathsf{P}_{\mathbf{Y}}}(\mathcal{Y}) \arrow[l, "U_{\mathbf{Y}}"] \end{tikzcd}$$

Frames

- $\mathcal{Y} = \{0, 1\}^k$, $\mathsf{P}_{\mathcal{Y}}$ probability measure on $\mathcal{Y}$.
- $\mathsf{P}_{\mathcal{Y}}$ may not be uniform probability measure!
- Functions χ_F, F a bit pattern form a spanning set of functions on $\mathcal{Y}$.
- $\{\chi_F\}$ form a (not necessarilly linearly independent) spanning set for $L^2_{\mathsf{P}_{\mathcal{Y}}}(\mathcal{Y})$.
  - By finite Fourier Analysis $\{\chi_F\}$ is an orthogonal basis in $L^2_{\mathsf{Uniform}_{\mathcal{Y}}}(\mathcal{Y})$.
- Any spanning set in a finite dimensional inner product space is a *frame*.
- Generalizes to group characters on finite abelian groups.

Matrices

- $\mathcal{Y}$, $\mathcal{Z}$ bitstring spaces.
- $T : L^2_{\mathsf{P}_{\mathbf{Z}}}(\mathcal{Z}) \to L^2_{\mathsf{P}_{\mathbf{Y}}}(\mathcal{Y})$ linear.
- Matrix of T:

  $$\mathsf{Matrix}(T) = \{\langle T\phi, \psi\rangle\} \quad \phi \in \mathsf{char}(\mathcal{Z}), \psi \in \mathsf{char}(\mathcal{Y})$$
- Differs from usual definition of matrix with respect to a basis.
- Compositionality:

  $$\mathsf{Matrix}(T \circ S) = \mathsf{Matrix}(T)\,\mathsf{Matrix}(S)$$

Examples

- $\Phi : \mathcal{Y} \to \mathcal{Z}$ is a homomorphism. Matrix(U_Φ) has only zero entries except for entries $\chi, \Phi_* \chi$ which are 1.
- Affine maps $\Psi : \mathcal{Y} \to \mathcal{Y}$. Matrix is diagonal, with diagonal entries having modulus 1.