# Математические и компьютерные основы защиты информации

Лекция 6



Антон Николаевич Гайдук | УНИВЕР vk.com/gaidukedu

16 марта 2023 г.

# Содержание дисциплины

#### Раздел I Введение

• Тема 1. Введение. История. Основные понятия.

#### Раздел II Симметричная криптография

- Тема 2 Классические шифры.
- Тема 3 Поточные алгоритмы шифрования.
- Тема 4 Блочные алгоритмы шифрования.
- Тема 5 Функции хэширования.
- Тема 6 Математические методы криптоанализа.

#### Раздел III Асимметричная криптография

- Тема 7 Протокол Диффи-Хэллмана.
- Тема 8 Криптосистемы с открытым ключом.
- Тема 9 Электронная цифровая подпись.

# Раздел II Симметричная криптография

#### Тема 6 Математические методы криптоанализа.

- Модель криптоаналитика
- Виды атак
- Параметры атак
- Статистические методы
- Линейный криптоанализ
- Дифференциальный криптоанализ
- Алгебраический криптоанализ
- Методы балансировки времени памяти

Математические методы криптоанализа

#### Модель криптоаналитика

- Криптоаналитику известна шифрсистема  $\mathcal{S}=\{\mathcal{K},\mathcal{P},\mathcal{C},E,D\}$ , неизвестен только используемый ключ  $k\in\mathcal{K}$ .
- Криптоаналитик наблюдает за *шифрсистемой* и получает полную или частичную информацию о парах «открытый текст-шифртекст»:

$$c_i = E_k(p_i), \quad p_i \in \mathcal{P}, c_i \in \mathcal{C}, \quad i = \overline{1, T},.$$

#### Задачи криптоанализа

- ullet Определить ключ k преобразования  $E_k$ ,
- ullet Не определяя ключ k, найти  $p_{T+1}$  по наблюдаемому  $c_{T+1} = E_k(p_{T+1}).$
- ullet Построить оценку  $\hat{k}$  ключа k преобразования  $E_k$  .
- ullet Не определяя ключ k, построить оценку  $\hat{p}_{T+1}$  по наблюдаемому  $c_{T+1}=E_k(p_{T+1}).$
- По парам «открытый текст-шифртекст»  $(p_i,c_i)$  определить, является ли преобразование  $e:c_i=e(p_i)$ . преобразованием зашифрования шифрсистемы  $\mathcal{S}:e\stackrel{?}{\in}E=\{E_k:k\in\mathcal{K}\}.$

#### Типы атак

Методы решения задач криптоанализа называются атаками.

- Известны  $c_i$  и некоторые свойства открытого текста  $p_i$  (атака при известном шифртексте),
- Известны  $p_i$  и  $c_i$  (атака при известном открытом тексте),
- Криптоаналитику представляется возможность выбирать заранее значения  $p_i$  (атака при выбранном открытом тексте),
- Криптоаналитик может задать значения  $p_i$  для  $i=2,\ldots,T$ , зная  $c_1,\ldots,c_{i-1}$  (атака при выбираемом открытом тексте),
- Криптоаналитик может выбирать шифртексты для расшифрования  $c_i$  (атака на основе подобранного шифртекста).
- Криптоаналитик знает взаимосвязи между различными ключами (атака на основе связанных ключей).

# Примеры информации об открытых текстах

Во время Второй мировой войны британская специальная служба обрабатывала шифрматериал немецкой шифровальной машины «Энигма». Для организации атаки при выбранном открытом тексте англичане по агентурным каналам доводили в немецкие подразделения информацию (ложную или правдивую) о наличии мин в тех или иных районах. Последующие сообщения немцев обязательно содержали слово «Minen» (мины, нем.).

Пакеты протокола HTTP, отправляемые сервером могут иметь фиксированные заголовки, за которыми следуют данные:

HTTP/1.1 200 OK

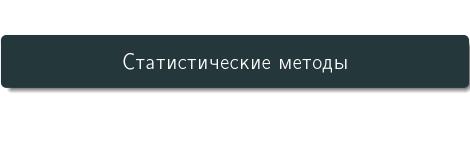
Server: Apache/2.2.10 (Unix) PHP/5.2.6 Content-Type: text/html; charset=UTF-8

# Примеры информации об открытых текстах

В сетях связи GSM второго поколения речевые данные оцифровываются. Каждым 18,4 мс разговора соответствует двоичное слово длиной 184. Для противодействия помехам в канале связи слово (как вектор-строка) умножается на двоичную матрицу размера 184 х 456. В результате получается кодовое слово X, которое обладает структурными особенностями: имеется 456 —184 независимых линейных комбинаций символов X, которые обязательно обращаются в 0 mod 2. Кодовое слово X разбивается на 4 фрейма - слова длиной 114. Каждый фрейм зашифровывается перед отправкой в канал связи. Криптоаналитик, который перехватывает зашифрованные фреймы, знает о стуктурных особенностях соответствующего открытого текста X (хотя не располагает информацией о самих речевых данных).

# Сложность методов криптоанализа характеризуется

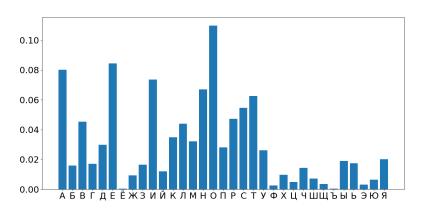
- Количеством T пар  $(p_i, c_i)$  (объем материала);,
- вычислительной сложностью атаки,
- объемом необходимой памяти для проведения атаки,
- вычислительной сложностью подготовительной стадии атаки,
- объемом необходимой памяти для подготовительной стадии атаки,
- вероятностью успеха атаки.



#### Частотные атаки

Естественные языки обладают статистическими закономерностями, которые может использовать криптоаналитик. Самые простые закономерности — разные частоты встречаемости символов, пар последовательных символов (биграмм), триграмм и т.д.

Шифр сдвига вщигь бгывг тывгозможное возможное в возможное возмо



# Частотные атаки: биграммы

В русском языке самая частая биграмма: СТ.

#### Шифр перестановки (длина блока: 4)

РАОП ДЬОЛ УТПС АНША ТСОЕ ТСРЙ ИНЦА ШВЕЕ ОТЙС РТОС ЕКТЧ СКТЕ

$$\sigma^{-1} = \left(\begin{array}{ccc} 1 & 2 & 3 & 4 \\ a & b & c & d \end{array}\right)$$

$$x_1x_2x_3x_4 \rightarrow x_ax_bx_cx_d$$

$$b = d + 1, a = b + 1, c = a + 1$$

$$\sigma^{-1} = \left( \begin{array}{ccc} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{array} \right)$$

ПАРОЛЬДОСТУПАНАШЕСТОЙСТРАНИЦЕВШЕСТОЙСТРОЧКЕТЕКСТ

# Тест $\chi^2$

Пусть  $X_1, X_2, \dots, X_n$  — н. о. р. CB со значениями в  $\{a_1, a_2, \dots, a_k\}$ . Рассмотрим две гипотезы:

$$H_0: X_1, X_2, \dots, X_n \sim F$$
  
 $H_1: X_1, X_2, \dots, X_n \nsim F$ 

Обозначим,  $\mathbb{P}_F\{X_j=a_i\}=p_i, \qquad (i=\overline{1,k},j=\overline{1,n}),$ 

$$u_i = \sum_{j=1}^n \mathbb{1}\{X_j = a_i\}, \qquad \mathbb{1} - \mathsf{индикатор} \ \mathsf{coбытия}$$

Вычисляем  $\chi^2$ -статистику Пирсона:

$$\chi^2 = \sum_{i=1}^k \frac{(\nu_i - np_i)^2}{np_i}.$$

# Тест $\chi^2$

 $\verb|wuymul| wcjbyl| nywbhckoy| cm u mcgjfy uhx yums gynbix iz yhwlsjncih nywbhckoy|$ 

Ключ	Открытый текст	Значение статистики $\chi^2$
1	vtxltk vbiaxk mxvagbjnx	77871.71
2	uswksj uahzwj lwuzfaimw	40605.19
3	trvjri tzgyvi kvtyezhlv	71357.46
4	squiqh syfxuh jusxdygku	55047.60
5	rpthpg rxewtg itrwcxfjt	32301.98
6	qosgof qwdvsf hsqvbweis	36089.68
7	pnrfne pvcure grpuavdhr	19274.26
8	omqemd oubtqd fqotzucgq	143531.85
9	nlpdlc ntaspc epnsytbfp	27060.57
10	mkockb mszrob domrxsaeo	35502.71
11	ljnbja lryqna cnlqwrzdn	50468.83
12	kimaiz kqxpmz bmkpvqycm	67245.95
13	jhlzhy jpwoly aljoupxbl	47829.83
14	igkygx iovnkx zkintowak	41474.49
15	hfjxfw hnumjw yjhmsnvzj	87055.24
16	geiwev gmtliv xiglrmuyi	22373.12
17	fdhvdu flskhu whfkqltxh	37453.80
18	ecguct ekrjgt vgejpkswg	32690.87
19	dbftbs djqifs ufdiojrvf	46897.27
20	caesar cipher technique	8858.44
21	bzdrzq bhogdq sdbgmhptd	51789.25
22	aycqyp agnfcp rcaflgosc	28163.64
23	zxbpxo zfmebo qbzekfnrb	80552.04
24	ywaown yeldan paydjemqa	31796.55
25	xvznvm xdkczm ozxcidlpz	158469.55

caesar cipher technique is a simple and easy method of encryption technique

### Пример

#### Определение (Шеннон)

Симметричная криптосистема называется совершенно криптостойкой, если апостериорное распределение вероятностей исходного случайного сообщения X при регистрации случайного шифртекста  $Y=E_k(X)$  совпадает с априорным распределением вероятностей:

$$\mathbb{P}\{X=x|Y=y\}=\mathbb{P}\{X=x\}, \qquad \forall x\in \mathcal{P}, y\in \mathcal{C}.$$

#### Следствие

Если выполняется условие совершенной криптостойкости, то количество информации по Шеннону, содержащейся в шифртексте Y об исходном сообщении X, равно нулю:

$$I(X,Y) = I(Y,X) = H(X) - H(X|Y) = H(Y) - H(Y|X) = 0.$$

#### Пример

Алфавит открытого текста:  $\{0,1\}$ .

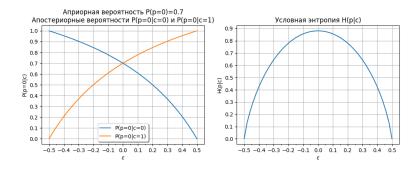
Бит открытого текста p имеет распределение:

$$\mathbb{P}{p=0} = 0.7, \qquad \mathbb{P}{p=1} = 0.3.$$

Выполняется зашифрование битом k с распределением:

$$\mathbb{P}\{k=0\} = 0.5 - \varepsilon, \qquad \mathbb{P}\{k=1\} = 0.5 + \varepsilon, \qquad \varepsilon \in [-0.5, 0.5].$$

$$c = p \oplus k$$
.



#### Линейный криптоанализ

#### Блочный алгоритм:

- 1. Количество раундов: r.
- 2. Раундовые ключи  $k_i, 1 \leqslant i \leqslant r$  незав. и равномерно распр. на  $\mathbb{B}^n$ .
- 3. Семействами раундовых функций зашифрования  $\hat{E}_{k_i}^{(i)}(P) = \sigma(P \oplus k_i)$ , где  $\sigma$  некоторая перестановка на множестве  $\mathbb{B}^n$ .

Определим скалярное произведение двух двоичных n-мерных векторов  $x=(x_1,x_2,\ldots,x_n)\in\mathbb{B}^n$ ,  $y=(y_1,y_2,\ldots,y_n)\in\mathbb{B}^n$  следующим образом:

$$x \cdot y = \bigoplus_{i=1}^{n} x_i y_i.$$

Для выполняемой на i-м раунде подстановки  $\sigma$  будем использовать линейную аппроксимацию  $(\alpha_{i-1},\alpha_i)\in\mathbb{B}^n imes\mathbb{B}^n$  со смещением:

$$\varepsilon_i = \mathbb{P}\{\alpha_{i-1} \cdot (C^{(i-1)} \oplus k_i) = \alpha_i \cdot \sigma(C^{(i-1)} \oplus k_i)\} - \frac{1}{2}, \quad \alpha_i \in \mathbb{B}^n, \quad i = \overline{1, r}.$$

#### Линейный криптоанализ

Последовательность линейных аппроксимаций  $\mathcal{A}=(\alpha_0,\dots,\alpha_r)$  называется r-раундовым линейным соотношением, значение  $(\alpha_0,\alpha_r)$  r-раундовой линейной аппроксимацией.

Обозначим

$$\mathcal{A}(k) = \alpha_0 \cdot k_1 \oplus \cdots \oplus \alpha_{r-1} \cdot k_r.$$

Смещение r-раундового линейного соотношения  $\mathcal{A}=(lpha_0,\dots,lpha_r)$ :

$$\varepsilon_{\mathcal{A}} = \mathbb{P}(\alpha_0 \cdot P \oplus \alpha_r \cdot E_k(P) = \mathcal{A}(k)) - \frac{1}{2}.$$

$$\varepsilon_{\mathcal{A}} = 2^{r-1} \prod_{i=1}^r \varepsilon_i.$$

Целью линейного криптоанализа является:

$$|\varepsilon_{\mathcal{A}}| \to \max_{\mathcal{A}=(\alpha_0,\dots,\alpha_r)}.$$

#### Линейный криптоанализ

$$\mathbb{P}\{k_5 \oplus k_7 \oplus k_{19} \oplus k_{25} = 0\} = \frac{1}{2} + \varepsilon,$$

Для определения  $k_5\oplus k_7\oplus k_{19}\oplus k_{25}$  требуется порядка  $\sim \frac{1}{arepsilon^2}$  уравнений.

# Дифференциаьлный криптоанализ

Пусть  $P_1$ ,  $P_2$  — два открытых текста, тогда разность (дифференциал)  $\Delta P = P_1 \oplus P_2 = \Delta C^{(0)}$ , порождает последовательность разностей  $\Delta C^{(i)} = C_1^{(i)} \oplus C_2^{(i)}, \Delta C^{(i)} \in \mathbb{B}^n$  промежуточных шифртекстов  $(1 \leqslant i \leqslant r)$  для каждого раунда i.

Последовательность разностей  $(\Delta P, \dots, \Delta C^{(r)})$  называется r-раундовым разностным соотношением, значение  $(\Delta P, \Delta C^{(r)})$  r-раундовой разностью. Вероятность разности для раунда i:

$$\mathbb{P}\{\Delta C^{(i)} = \delta_i | \Delta C^{(i-1)} = \delta_{i-1}\} = p_{\delta_i, \delta_{i-1}}, \delta_0, \dots, \delta_r \in \mathbb{B}^n, i = \overline{1, r},$$

Вероятность r-раундового разностного соотношения:

$$\mathbb{P}(\Delta C^{(1)} = \delta_1, \dots, \Delta C^{(r)} = \delta_r | \Delta P = \delta_0) = \prod_{i=1}^{n_r} p_{\delta_i, \delta_{i-1}} \to \max_{\Delta C^{(1)}, \dots, \Delta C^{(r)}}.$$

# Алгебраические методы

# Алгебраический криптоанализ

Алгебраическая атака на блочный алгоритм шифрования, основанная на решении системы полиномиальных уравнений состоит из следующих шагов:

- нахождение системы соотношений, описывающих блочный алгоритм шифрования, при известных парах открытого и зашифрованного текстов. Неизвестные переменными являются биты(байты) секретного ключа;
- решение полученной системы соотношений с использованием соответствующего алгоритма (методы линеаризации, методы основанные на базисах Гребнера, SAT решатели).

### Алгебраический криптоанализ

Существует много способов алгебраического описания блочного алгоритма шифрования. Для криптоаналитика интерес представляют такие алгебраические соотношения решение которых имеют наименьшую вычислительную сложность.

#### Алгебраический криптоанализ

Пусть имеется d+1 пара  $(p_i,c_i),\in\mathbb{B}^n imes\mathbb{B}^n$  открытый текст/шифртекст зашифрованных на одном ключе, тогда интерполяционная формула Лагранжа утверждает, что функция  $f:\mathbb{B}^n\to\mathbb{B}^n$  отображающая  $p_i$  в  $c_i$  имеет следующий вид:

$$f(x) = \sum_{i=0}^{d} c_i \prod_{\substack{j=0, \ j \neq i}}^{d} \left( \frac{x - p_j}{p_i - p_j} \right).$$

Если итерационный блочный алгоритм шифрования с длиной блока n может быть представлен как полином степени  $d < 2^n$  с коэффициентами зависящими от пар открытый текст/шифртекст (при фиксированном ключе), тогда существует интерполяционная атака имеющая сложность  $O(d\log d)$ .

# Система полиномиальных уравнений

Пусть задано некоторое конечное поле  $\mathbb{K}.$ 

Рассмотрим вектор  $\alpha=(\alpha_1,\alpha_2,\dots,\alpha_n)$ , у которого каждая координата  $\alpha_i,i=\overline{1,n}$  целое неотрицательное число. Тогда мономом от переменных  $x_1,x_2,\dots,x_n$  будем называть произведение вида:

$$x^{\alpha} = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}.$$

Отметим, что если  $\alpha=(0,0,\dots,0)$ , то  $x^\alpha=1$ . Через  $|\alpha|=\alpha_1+\alpha_2\dots\alpha_n$  будем обозначать полную степень монома  $x^\alpha$ .

Полином f от переменных  $x_1,x_2,\ldots,x_n$  с коэффициентами из  $\mathbb K$  будем называть конечную линейную комбинацию  $f(x_1,x_2,\ldots,x_n)=\sum_{\alpha}a_{\alpha}x^{\alpha}$  мономов  $x^{\alpha}$  с коэффициентами  $a_{\alpha}$  из поля  $\mathbb K$ . Полной степенью полинома f будем называть величину равную  $max_{\alpha}|\alpha|$ .

И пусть заданы полиномы  $f_1, f_2, \dots f_m \in \mathbb{K}[x_1, x_2, \dots, x_n]$ . Рассмотрим следующую систему уравнений:

$$S: f_i(x_1, x_2, \dots, x_n) = 0, i = \overline{1, m}.$$
 (1)

#### Базовый метод линеаризации

Метод линеаризации является хорошо известным методом решения больших систем многомерных полиномиальных уравнений. Суть метода заключается в том, что нелинейная система уравнений приводится к линейной путем обозначения каждого монома степени большей 1 через новую независимую переменную, т.е. моном  $x^{\alpha}$  для которого  $|\alpha|>1$  заменяется на новую переменную  $y_j$ . После этого, система уравнений решается методом Гаусса. Заметим, что решение системы содержит как изначальные переменные, так и новые переменные. Например, мономы

$$x_i^2, x_j^2, x_i x_j$$

были заменены на переменные

$$y_{ii}, y_{jj}, y_{ij},$$

Эффективность метода линеаризации зависит от количества линейно независимых уравнений. Для того чтобы применить метод линеаризации, число линейно независимых уравнений в системе должно быть примерно таким же, как число мономов.

Например, для решения уравнений степени d над кольцом  $\mathbb{F}_2[x_1,\dots,x_n]$  методом линеаризации, потребуется не более чем  $\sum_{i=1}^d \binom{n}{i}$  переменных. Поэтому если число линейно независимых уравнений больше либо равно $\sum_{i=1}^d \binom{n}{i}$  тогда можно ожидать, что решение будет найдено, иначе метод линеаризации может не сработать.

Методы балансировки времени памяти Time Memory tradeoff

# Инвертирование однонаправленных отображений

Задача определения ключа алгоритма шифрования сводится к инвертированию некоторого отображения

$$F:X\to Y,\quad F(x)=y,\quad x\in X,\quad y\in Y,$$

где X, Y — конечные множества, |X|=N.

Дано: отображение  $F:X \to Y$ , образ  $y \in Y$ . Найти: прообраз  $x \in X$  такой, что F(x) = y.

#### Пример

$$F = E_p : \mathcal{K} \to \mathcal{C},$$
  

$$E_p(k) = E_k(p) = c,$$
  

$$k \in \mathcal{K}, \quad p \in \mathcal{P}, \quad c \in \mathcal{C}.$$

# Методы балансировки времени-памяти (ТМТО)

#### Основные

- Метод Хеллмана (1980 г.)
- Метод особых точек (1982 г.)
- Метод радужных таблиц (2003 г.)

# Параметры и характеристики ТМТО

#### Основные параметры

l — количество таблиц;

m — количество цепочек (строк) в одной таблице;

t — длина одной цепочки (количество столбцов в таблице);

 $R_j^{(k)}:Y o X$  — редуцирующее отображение;

 $F_j^{(k)}: X o X$  ,  $F_j^{(k)} = R_j^{(k)} \circ F$  — инвертируемое преобразование.

#### Основные характеристики

Q — трудоемкость предварительного этапа по времени;

T — трудоемкость оперативного этапа по времени;

M — трудоемкость по памяти;

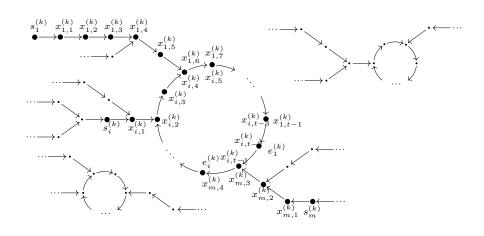
P — вероятность успеха.

# Метод Хеллмана

#### Схема построения k-й таблицы

$$\begin{pmatrix} x & \stackrel{F}{\rightarrow} \end{pmatrix} y \stackrel{R^{(k)}}{\rightarrow} & \tilde{x}_0^{(k)} & \stackrel{F^{(k)}}{\rightarrow} & \tilde{x}_1^{(k)} & \stackrel{F^{(k)}}{\rightarrow} & \dots & \stackrel{F^{(k)}}{\rightarrow} & \tilde{x}_j^{(k)} = e_i^{(k)} \\ \parallel? & & \parallel \\ s_i^{(k)} = x_{i,0}^{(k)} & \stackrel{F^{(k)}}{\rightarrow} & \dots & \stackrel{F^{(k)}}{\rightarrow} & x_{i,t-j-1}^{(k)} & \stackrel{F}{\rightarrow} & x_{i,t-j}^{(k)} \\ \end{pmatrix}$$

# Метод Хеллмана



#### Метод Хеллмана

#### Теорема

Пусть преобразование  $F^{(k)}: X \to X$  является случайным, то есть выбирается равновероятно из множества мощности  $N^N$  всех отображений X в себя. Пусть прообраз x выбирается из X равновероятно. Тогда

$$P^{(k)} \geqslant \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{t} \left(1 - \frac{it}{N}\right)^{j}.$$

# Метод особых точек и радужных таблиц

#### Метод особых точек

$$s_{1}^{(k)} = x_{1,0}^{(k)} \xrightarrow{F^{(k)}} x_{1,1}^{(k)} \xrightarrow{F^{(k)}} \dots \xrightarrow{F^{(k)}} x_{1,t_{1-1}}^{(k)} \xrightarrow{F^{(k)}} x_{1,t_{1}}^{(k)} = e_{1}^{(k)} \in E$$

$$\vdots$$

$$s_{i}^{(k)} = x_{i,0}^{(k)} \xrightarrow{F^{(k)}} x_{i,1}^{(k)} \xrightarrow{F^{(k)}} \dots \xrightarrow{F^{(k)}} \dots \xrightarrow{F^{(k)}} \dots \xrightarrow{F^{(k)}} x_{i,t_{i-1}}^{(k)} \xrightarrow{F^{(k)}} x_{i,t_{i}}^{(k)} = e_{i}^{(k)} \in E$$

$$\vdots$$

$$s_{m}^{(k)} = x_{m,0}^{(k)} \xrightarrow{F^{(k)}} x_{m,1}^{(k)} \xrightarrow{F^{(k)}} \dots \xrightarrow{F^{(k)}} \dots \xrightarrow{F^{(k)}} x_{m,t_{m-1}}^{(k)} \xrightarrow{F^{(k)}} x_{m,t_{m}}^{(k)} = e_{m}^{(k)} \in E$$

#### Метод радужных таблиц

$$s_{1}^{(k)} = x_{1,0}^{(k)} \xrightarrow{F_{1}^{(k)}} x_{1,1}^{(k)} \xrightarrow{F_{2}^{(k)}} x_{1,2}^{(k)} \xrightarrow{F_{3}^{(k)}} \dots \xrightarrow{F_{t-1}^{(k)}} x_{1,t-1}^{(k)} \xrightarrow{F_{t}^{(k)}} x_{1,t}^{(k)} = e_{1}^{(k)}$$

$$\vdots$$

$$s_{i}^{(k)} = x_{i,0}^{(k)} \xrightarrow{F_{1}^{(k)}} x_{i,1}^{(k)} \xrightarrow{F_{2}^{(k)}} x_{i,2}^{(k)} \xrightarrow{F_{3}^{(k)}} \dots \xrightarrow{F_{t-1}^{(k)}} x_{i,t-1}^{(k)} \xrightarrow{F_{t}^{(k)}} x_{i,t}^{(k)} = e_{i}^{(k)}$$

$$\vdots$$

$$s_{m}^{(k)} = x_{m,0}^{(k)} \xrightarrow{F_{1}^{(k)}} x_{m,1}^{(k)} \xrightarrow{F_{2}^{(k)}} x_{m,2}^{(k)} \xrightarrow{F_{3}^{(k)}} \dots \xrightarrow{F_{t-1}^{(k)}} x_{m,t-1}^{(k)} \xrightarrow{F_{t}^{(k)}} x_{m,t}^{(k)} = e_{m}^{(k)}$$

