Курс лекций по теоретической криптографии

Тема 2. **Основные положения теории Шеннона**

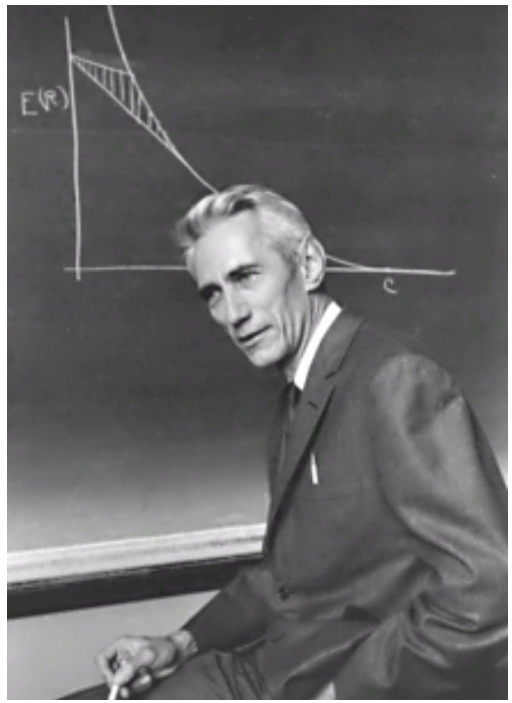
Шокуров А.В.

Клод Элвуд Шеннон 1916 — 2001

A Mathematical Theory of Cryptography.
Memorandum MM 45-110-02, Sept. 1, 1945.
Bell Laboratories (classified report).

A Mathematical Theory of Communication //
Bell System Technical Journal. 1948.
Vol. 27 (3). Pp. 379–423.
——————————— Vol. 27 (4).
Pp. 623–656.

Communication Theory of Secrecy Systems
Bell System
Technical Journal. 1949. Vol. 28 (4).
Pp. 656–715.



Перевод на русский язык в книге:
К. Шеннон. Работы по теории информации и кибернетике. — М.: Иностранная литература, 1963. — Сс. 333–369.

Общие предположения

- Информация дискретна. Сообщение — слово в конечном алфавите (последовательность символов этого алфавита).
- Язык — вероятностный процесс, создающий сообщения. Соответственно, каждое сообщение имеет определённую вероятность, то есть на пространстве сообщений задано некоторое распределение вероятностей.
- Шифр — множество взаимно однозначных отображений множества возможных сообщений в множество криптограмм. Конкретное отображение соответствует шифрованию на конкретном ключе, т.е.ключ однозначно определяет отображение.

Общие предположения

- Противник не ограничен[1] в ресурсах (вычислительных, временны́х), но имеет доступ только к передаваемым по каналу связи криптограммам.
- Противник может знать, какой шифр и язык использовались, но не знает, какой ключ был применён при получении данной криптограммы.

[1] Особенность теоретико-информационного подхода — в отличие от теоретико-сложностного.

Примеры: Шифры замены

A, B — алфавиты

$m = m_0 m_1 m_2 \ldots m_{l-1}$ — сообщение, $\quad m_i \in A$

$\varphi : A \to B$ — инъективное[2] отображение

$$c = c_0 c_1 c_2 \ldots c_{l-1} = \varphi(m_0)\varphi(m_1)\varphi(m_2) \ldots \varphi(m_{l-1}), \quad c_i \in B$$

- Шифр Цезаря (на современном латинском алфавите из 26 букв):

$$A = B \leftrightarrow \{0, 1, 2, \ldots, 25\}$$
$$\varphi(m_i) = m_i + 3 \mod 26$$

- Обобщение шифра Цезаря:

$$A = B \leftrightarrow \{0, 1, 2, \ldots, 25\}, \quad -25 \leqslant k \leqslant 25 \text{ — ключ}$$
$$\varphi_k(m_i) = m_i + k \mod 26$$

- Шифр Виженера с секретным (шифрующим) словом длины s

$$A = B \leftrightarrow \{0, 1, 2, \ldots, 25\}, \quad k = (k_0, k_1, \ldots, k_{s-1}) \in A^s$$ — ключ

$$\varphi_k(m_i) = m_i + k_{i \bmod s} \mod 26$$

[2] Инъективность: $\forall x, y \quad \big(x \neq y \Rightarrow \varphi(x) \neq \varphi(y)\big)$.

Примеры: Шифры перестановки (с периодом d)

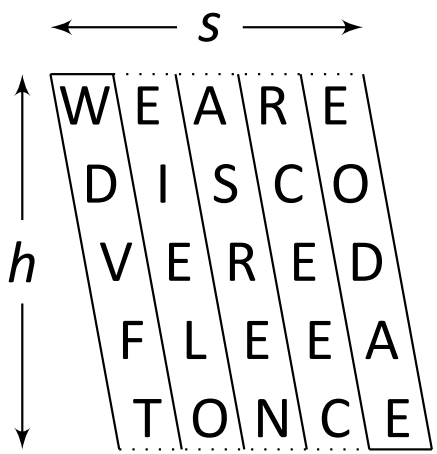
$m = m_1 m_2 m_3 \ldots m_l$ — сообщение

$\sigma \in S_d$ — перестановка на d элементах (считаем, что $l = t \cdot d$)

$$c = c_1 c_2 c_3 \ldots c_l = m_{\sigma(1)} m_{\sigma(2)} \ldots m_{\sigma(d)} m_{\sigma(1)+d} \ldots m_{\sigma(d)+(t-1)\cdot d}$$

- Шифр сциталы при длине окружности жезла h (в буквах), $d = l = s \cdot h$:

$$\sigma(i) = s \cdot (i-1) + \left\lfloor \frac{i-1}{h} \right\rfloor + 1$$



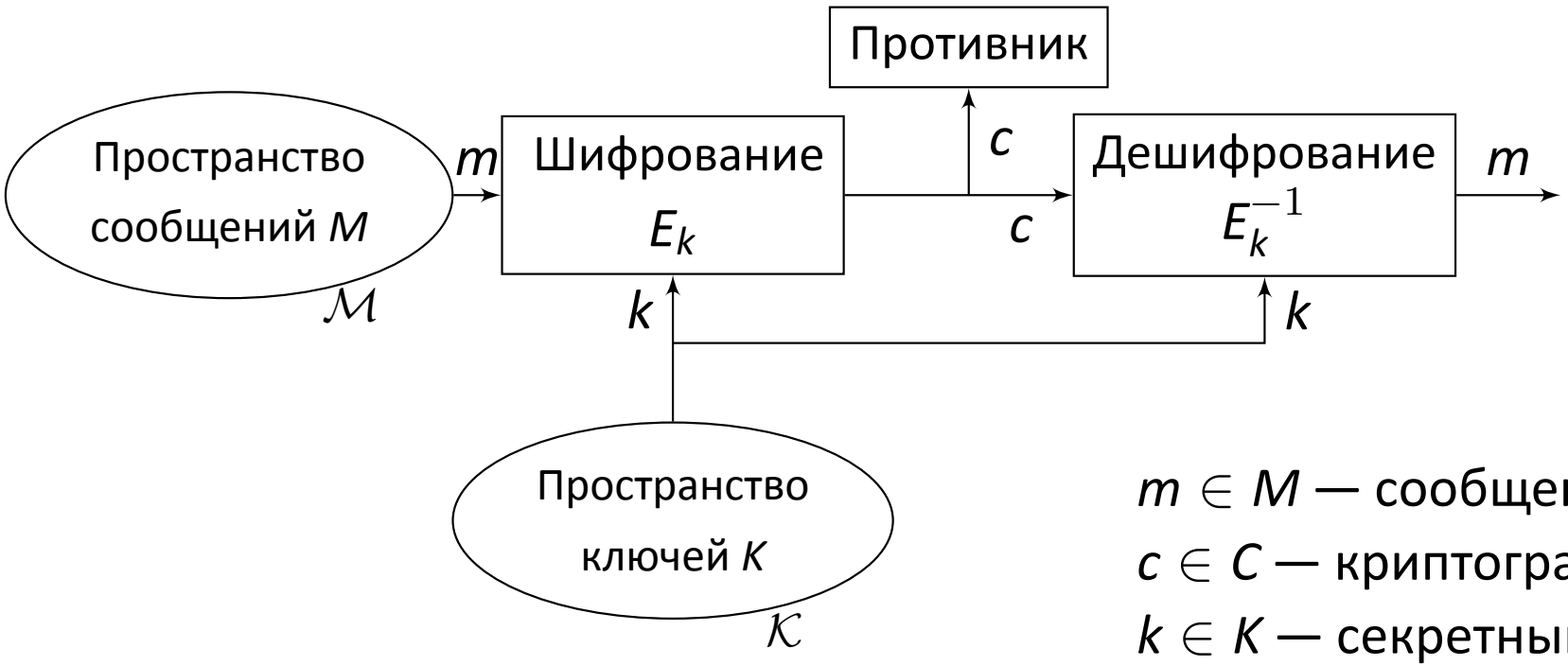
$m =$ WEAREDISCOVEREDFLEEATONCE

$\downarrow$

$c =$ WDVFTEIELOASRENRCEECEODAE

? Проверьте, верно ли выписана формула шифрующего преобразования. Действительно ли определённая тут σ является перестановкой, то есть взаимно однозначным отображением множества $\{1, 2, \ldots, l\}$ на себя?

Общая схема



$m \in M$ — сообщение (ОТ)

$c \in C$ — криптограмма (ШТ)

$k \in K$ — секретный ключ

Система связи со свойством секретности (СССС)

Определение

Система связи с секретностью (secrecy system) S — семейство однозначно обратимых (то есть инъективных) отображений $\{E_k : M \to C\}_{k \in K}$, где M, K, C — конечные непустые множества (пространства сообщений, ключей и криптограмм), причём на M и K заданы распределения вероятностей $\mathcal{M}$ и $\mathcal{K}$ соответственно.

Это — одна из математических моделей шифра с секретным ключом. $\mathcal{K}$ задаёт вероятности отображений E_k из S. Обратимость отображений шифрования необходима для однозначного дешифрования криптограммы.

Система связи со свойством секретности (СССС)

Задача противника — извлечь m или k (*угроза*) из перехваченной криптограммы $c = E_k(m)$ (*атака*), где $m \leftarrow \mathcal{M},\ k \leftarrow \mathcal{K}$, зная $S, \mathcal{M}, \mathcal{K}$.

$\mathcal{M}$ — **априорное** (до атаки) распределение вероятностей на M «по мнению» противника[3].

$\mathcal{M}_c$ — **апостериорное** (после атаки) распределение вероятностей на M с точки зрения противника при перехвате c (условное распределение вероятностей).

То же — для распределения вероятностей на пространстве ключей.

[3] Например, $\Pr[m]$ — относительная частота появления последовательности m в нормативном английском языке.

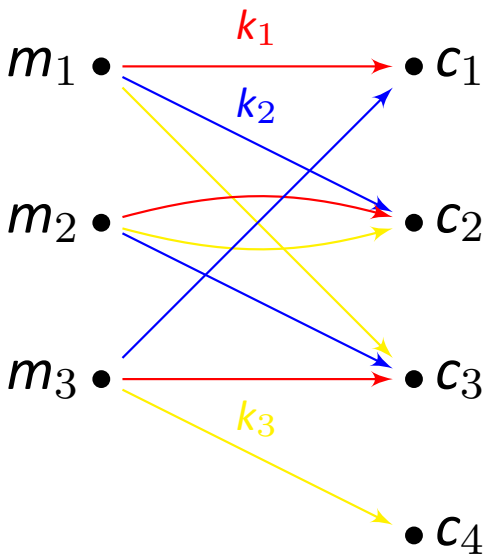
Свойства систем

- **Замкнутая** система: для всех $k \in K$ отображения E_k сюръективны.

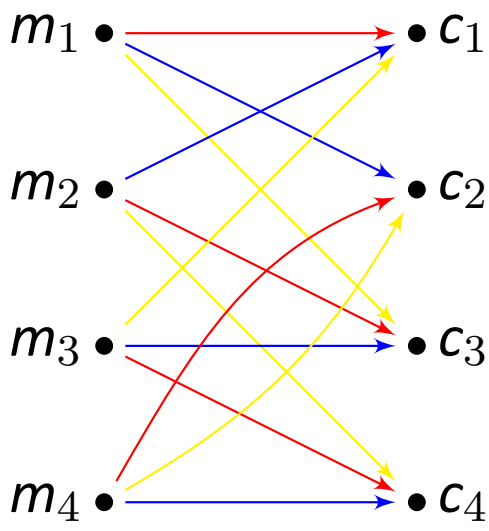
  Каждая криптограмма может быть получена на всех возможных ключах.
- **Чистая** система: для любых $k_1, k_2, k_3 \in K$ найдётся такой k_4, что $E_{k_1} E_{k_2}^{-1} E_{k_3} = E_{k_4}$ и все ключи равновероятны. Апостериорные вероятности сообщений не зависят от выбора ключа. Например, в чистом шифре простой замены важен только характер повторений букв, сами буквы — несущественная маскировка.
- **Смешанная** система — не являющаяся чистой.
- **Подобные** системы: $S_1 \approx S_2$, если существует обратимое отображение U, такое, что $S_1 = US_2$. Подобные системы эквивалентны с точки зрения противника. Между их пространствами криптограмм можно установить взаимно однозначное соответствие. Достаточно рассматривать один тип систем их множества подобных. Например, шифр Цезаря и обратный шифр Цезаря подобны, U — обращение алфавита в криптограмме.

Графическое представление систем

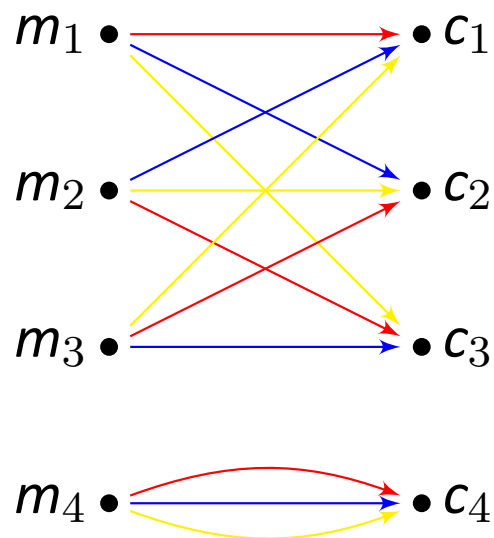
Система связи с секретностью (при $|K| = 3$):



произвольная



замкнутая



чистая

Совершенная секретность

Задача — формализовать следующее свойство системы:

перехват криптограммы не даёт противнику никакой новой (дополнительной) информации о зашифрованном сообщении.

$\Pr[m]$ — априорная вероятность сообщения m в соответствии с вероятностным распределением $\mathcal{M} = \{(m, \Pr[m])\}_{m \in M}$

$\Pr[m|c]$ — апостериорная вероятность сообщения m при условии, что противник перехватил криптограмму $c \in C = \bigcup\limits_{k \in K} E_k(M)$

Определение

Система $S = \{E_k\}_{k \in K}$ *обеспечивает совершенную секретность*, или является *совершенной*, если при любой перехваченной криптограмме $c \in C$ апостериорное распределение вероятностей на M совпадает с априорным, то есть по всем $m \in M$ справедливо равенство

$$\Pr[m|c] = \Pr[m].$$

Необходимые условия совершенной секретности

Теорема

$S = \{E_k\}_{k \in K}$ совершенна $\Leftrightarrow$ для всех $c \in C$ и $m \in M$ $\Pr[c|m] = \Pr[c]$.

Доказательство. По формуле Байеса

$$\Pr[m|c] = \frac{\Pr[m] \cdot \Pr[c|m]}{\Pr[c]}.$$

Следовательно, либо $\Pr[m] = 0$ для всех m, что невозможно, либо $\Pr[c] = \Pr[c|m]$. Очевидно, верно и обратное. □

Необходимые условия совершенной секретности

$\Pr[c|m]$ — сумма вероятностей всех тех ключей k, для которых $E_k(m) = c$. В совершенных системах $\Pr[c|m]$ не зависит от m. Таким образом, полная вероятность всех ключей, переводящих m в c,равна полной вероятности всех ключей, переводящих m' в c, для любого m'.

При фиксированном k каждое E_k — инъективное отображение M в C $\Rightarrow \boxed{|C| \geqslant |M|}$

При фиксированном m для любой $c \in C$ $\Pr[c|m] = \Pr[c] \neq 0$.

Значит, для любой c существует $k \in K$, такой, что $E_k(m) = c$, причём для разных криптограмм ключи, переводящие в них m, должны быть разными $\Rightarrow \boxed{|K| \geqslant |C|}$

Необходимые условия совершенной секретности

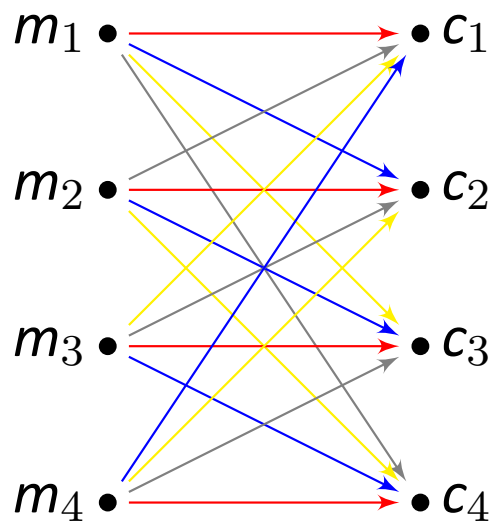
Из этих и других соображений получаем следующие условия.

Число различных ключей должно быть не меньше числа возможных сообщений

Все ключи равновероятны ($\mathcal{K}$ — равномерное распределение)

Каждый ключ используется не более одного раза

Можно построить совершенную систему при $|K| = |M|$



Представление совершенных систем — полные двудольные графы

Пример совершенной системы: Шифр Вернама

Шифр Вернама со случайным одноразовым ключом той же длины, что и сообщение, — пример совершенной системы связи с секретностью:

$$
\begin{aligned}
m &= \texttt{10010111010111100011011110}\ldots\texttt{1} \quad m \in M = \{0,1\}^n, \quad |m| = n \\
\oplus \quad & \\
k &= \texttt{00011011111110011101111100}\ldots\texttt{0} \quad k \in_R K = \{0,1\}^n, \quad |k| = n, \\
& \downarrow \\
c &= \texttt{10001100101001111110100010}\ldots\texttt{1} \quad c \in \{0,1\}^n, \quad |c| = n
\end{aligned}
$$

Пример совершенной системы: Шифр Вернама

Почему сложение с чисто случайным битом даёт чисто случайный бит?

$$\begin{array}{rll} \beta \in_R \{0,1\}: & \Pr[\beta = 0] = p_0 \\ & \Pr[\beta = 1] = p_1 \\ & p_0 + p_1 = 1 \\ \alpha \in_R \{0,1\}: & \Pr[\alpha = 0] = \frac{1}{2} \\ & \Pr[\alpha = 1] = \frac{1}{2} \end{array}$$

$$\Rightarrow \quad \beta \oplus \alpha = \begin{bmatrix} 0 \text{ с вер. } \frac{1}{2}p_0 \\ 1 \text{ с вер. } \frac{1}{2}p_0 \\ 1 \text{ с вер. } \frac{1}{2}p_1 \\ 0 \text{ с вер. } \frac{1}{2}p_1 \end{bmatrix} = \begin{bmatrix} 0 \text{ с вер. } \frac{1}{2} \\ 1 \text{ с вер. } \frac{1}{2} \end{bmatrix}$$

Оценка качества системы

Не все системы совершенны. Как оценивать используемые на практике шифры?

Критерии оценки:

- «количество секретности» ($\approx$ стойкость) — с какими затратами времени и сил и с каким объёмом материала противник достигает успеха $\uparrow$
- объём (длина) ключа $\downarrow$
- трудоёмкость шифрования / дешифрования $\downarrow$
- разрастание числа ошибок в процессе вычислений и передачи данных $\downarrow$
- увеличение длины криптограммы по сравнению с исходным сообщением $\downarrow$

Теория говорит, что улучшить все характеристики одновременно невозможно.

Например, нельзя одновременно повышать «количество секретности» и уменьшать длину ключей, как следует свойств совершенных систем.

Энтропия как мера неопределённости

Количество информации, создаваемой при выборе сообщения, можно измерять с помощью **энтропии**:

$$\left.\begin{aligned} H_M &= -\sum_{m\in M} \Pr[m] \log \Pr[m] \\ H_K &= -\sum_{k\in K} \Pr[k] \log \Pr[k] \end{aligned}\right\} \text{— мера } \textbf{неопределённости}$$

сообщения / ключа

$$0 \leqslant H_M \leqslant \log |M|$$

0: при $|M| = 1$

$\log |M|$: при равновероятном выборе сообщений

Для определения теоретической меры секретности используется **условная энтропия**:

$$\left.\begin{aligned} H_{M|C}(n) &= -\sum_{\substack{c\in C_n \\ m\in M}} \Pr[m,c] \log \Pr[m|c] \\ H_{K|C}(n) &= -\sum_{\substack{c\in C_n \\ k\in K}} \Pr[k,c] \log \Pr[k|c] \end{aligned}\right\} \text{— мера } \textbf{ненадёжности}$$

сообщения / ключа

(C_n — подмножество криптограмм длины n)

Ещё некоторые понятия

Расстояние единственности — минимальное число букв «средней криптограммы», позволяющее однозначно восстановить исходное сообщение ($H_{M|C}(n) = 0$)

Идеальная система — та, у которой $H_{M|C}(n)$ и $H_{K|C}(n)$ не стремятся к нулю при $n \to \infty$ (бесконечное расстояние единственности)

Строго идеальная система — та, у которой $H_{K|C}(n) = H_K$ для всех n

Избыточность языка измеряет, на сколько может быть уменьшена длина текста без потери информации

Рабочая характеристика шифра — функция от длины криптограммы, равная среднему объёму работы, затрачиваемой на нахождение ключа

Использование при построении (блоковых) шифров нескольких циклов **перемешивания** (confusion) [$\sim$ замена] и **рассеивания** (diffusion) [$\sim$ перестановка]

Дополнение. Группы

Определение

Множество G с операцией композиции, которая сопоставлет каждой паре $a, b \in G$ третий элемент $c \in G$, обозначаемой через $a \cdot b$ или просто ab.Для этой операции выполняются условия

1. Закон ассоциативности: для любых $a, b, c \in G$ выполняется равенство $a(bc) = (ab)c$.
2. Существует левая единица $e \in G$ — такой элемент,что для любого $a \in G$ всегда $ea = a$.
3. Существует левый обратный: для любого $a \in G$ существует такой $b \in G$, что $ba = e$.

Лемма

Если $ba = e$, то $ab = e$. Левая единица является правой единицей, т.е. для всех $a \in G$ выполняется $ae = a$.

Дополнение. Группы

Доказательство.

Пусть $ba = e$. Тогда $(ba)b = eb = b$. По аксиоме 1 для элемента b существует такой c, что $cb = 1$. Тогда

$$e = cb = c((ba)b) = (cb)ab = eab = ab.$$

Далее, согласно доказанному,

$$ae = a(ba) = (ab)a = ea = a.$$

□

Следствие

Уравнения $ax = b$ и $ya = b$ имеют решения.

Следствие

Из $ax = ax'$ и $ya = y'a$ следует, что $x = x'$ и $y = y'$.

Дополнение. Группы

Следствие

В группе существует все левые и правые единицы единственны. Левая и правая единицы равны, все обратные левые и правые для элемента a равны.

В силу последней леммы обратный элемент к a будем обозначать через a^{-1}, а единицу через 1.

Определение

Подстановкой множества M называется взаимно однозначное отображение этого множества на себя. Если это множество конечно и его элементы занумерованы числами $1, 2, \ldots, n$, то каждую подстановку можно описать таблицей, в которой под каждым числом указывается его номер его образа. Например,

$$s = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$$

изображает подстановку чисел 1, 2, 3, 4, в которой 1 переходит в 2, 2 переходит в 4, 3 переходит в 3, 4 переходит в 1. Множество подстановок на множестве из n элементов обозначается через S_n. Это множество является группой относительно композиции отображений.

Задачи

Задача 1. Докажите, что множество S_n является группой относительно композиции отображений.

Задача 2. Проверьте, верно ли выписана формула шифрующего преобразования на стр. 6. Действительно ли определённая тут σ является перестановкой, то есть взаимно однозначным отображением множества $\{1, 2, \ldots, l\}$ на себя?

Задача 3. Проверьте ”чистоту”системы шифрования в третьем столбце на стр. 11.