УДК 004.1

А.В. Кизино, маг.; О.А. Новосельская, доц.
(БГТУ, г. Минск)

ИНДЕКСИРОВАНИЕ СИМВОЛОВ КАК МЕТОД ИЗМЕНЕНИЯ ВЕРОЯТНОСТНОЙ СТАТИСТИКИ В ПОДСТАНОВОЧНЫХ ШИФРАХ

Вероятностная статистика – это раздел математики, который изучает закономерности случайных явлений. В лингвистике и обработке информации фундаментальным свойством является статистическая устойчивость. Согласно закону больших чисел (теореме Бернулли), при бесконечном увеличении длины текста N относительная частота появления символа всегда стремится к его истинной вероятности $P(a)$. Этот предел частоты выступает уникальной характеристикой, своеобразным «отпечатком», любого естественного языка. Естественные языки отличаются высокой степенью избыточности и крайне неравномерным распределением частот символов, что описывается законом Ципфа [1]. Данный закон гласит, что частота символа обратно пропорциональна его рангу. Например, в русском языке самой частой гласной является «О» с вероятностью около 10.9%. Буква «Е» встречается с частотой около 8.4%. Редкие буквы, такие как «Ф», «Щ», «Ъ», имеют частоту менее 1%. На гистограмме такое распределение выражается в наличии ярко выраженных «пиков» высокочастотных символов и «хвостов» редких символов (рис.1).

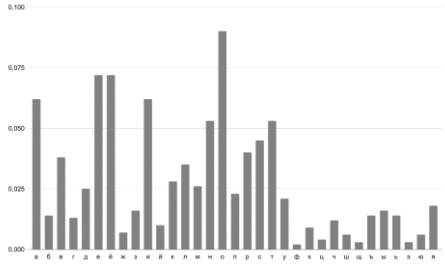

Рисунок 1 – Гистограмма распределения символов по частоте появления

Именно эти пики позволяют криптоаналитикам идентифицировать язык и буквы даже в зашифрованных сообщениях. В симметричных криптосистемах иногда применяются шифры простой замены (моноалфавитные подстановки), где каждой букве открытого текста ставится в соответствие уникальный символ шифротекста. Основная проблема таких шифров кроется в том, что операция замены является изоморфизмом относительно вероятностного распределения. Если в открытом тексте буква «О» встречается с вероятностью 10.9%, то и её заменитель в шифротексте будет иметь точно такую же вероятность.

Используя метод частотного анализа, криптоаналитик строит гистограмму шифротекста, сравнивает её с эталонной гистограммой языка и сопоставляет самые высокие «пики». Из-за этой уязвимости простые подстановочные шифры абсолютно ненадежны. Даже полиалфавитные системы (например, шифр Виженера) оставляют периодические следы, которые можно выявить методом Касиски [2].

Для решения проблемы сохранения частотного рельефа предлагается метод вероятностного разбиения. Идея базируется на отказе от отображения «один к одному» (1:1) в пользу отображения «один ко многим» (1:M) со случайным выбором конкретного варианта.

Алгоритм преобразования состоит в следующем. Вводится расширенный алфавит, состоящий из биграмм – пар исходной буквы и случайно выделенного индекса. Из входного потока читается символ и генерируется случайное целое число в заданном диапазоне с не равномерным распределением. Формируется новый символ-пара. Например, слово «СЛОВО» может быть преобразовано в последовательность «С1, Л0, О0, В2, О1». С точки зрения статистики, символы «О0» и «О1» теперь являются разными элементами расширенного алфавита. Вероятность появления такой пары определяется как произведение исходной вероятности буквы на вероятность выбора индекса (1).

$$P\left(Y=\left(x,k\right)\right)=P\left(X=x\right)\cdot P\left(Index=k\mid X=x\right)=P\left(x\right)\cdot\frac{1}{M},\qquad(1)$$

где $P(X=c)$ – исходная вероятность появления буквы «х» в языке.

Таким образом, исходный частотный пик буквы разбивается на более мелкие пики, высота каждого из которых уменьшается пропорционально количеству индексов. На конечных выборках количество раз, когда букве будет присвоен конкретный индекс, подчиняется биномиальному распределению. Стандартное отклонение количества появлений конкретной пары (a,k) составляет:

$$\sigma\approx\sqrt{N_a\cdot\frac{1}{M}\cdot\left(1-\frac{1}{M}\right)}.\qquad(2)$$

Это означает, что для редких букв возможны ситуации, когда некоторые индексы вообще не выпадут, что создает дополнительный «шум» для криптоаналитика.

В рамках исследования метод был протестирован на фрагменте текста на русском языке длиной 310 символов. Использовался диапазон из трех случайных индексов {0, 1, 2} Исходный доминирующий символ «Е» (имевший частоту 10.6%) распался на несколько подгрупп. Самая частая новая вариация «Е1» получила частоту всего

3.8%, что сопоставимо с частотой редкой исходной буквы «А». Максимальная вероятность символа в тексте упала более чем в 2 раза: с 0.1765 до 0.0784. Анализ гистограмм (рис. 2) четко показывает изменение формы распределения.

Исходный график с высокими пиками растянулся по оси X, так как количество категорий увеличилось в 3 раза. Вместо резких перепадов образовалось «плато» низких вероятностей. Это «размазывает» статистику и визуально превращает распределение в шумоподобное. Кроме того, анализ внутреннего распределения индексов показал, что для коротких выборок характерна сильная асимметрия. Этот эффект недетерминированности крайне полезен: криптоаналитик не может гарантировать равномерное представление каждого варианта буквы, что усложняет построение статистических моделей.

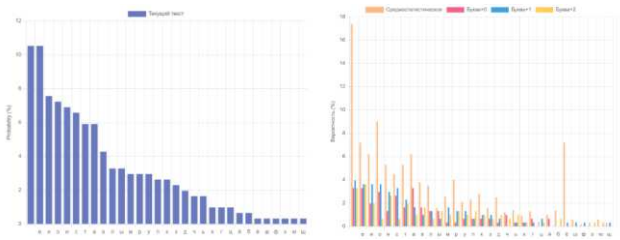

Рисунок 2 – Гистограмма распределения символов до и после обработки

Представленный метод случайного индексирования является эффективным средством предобработки текста перед его шифрованием. Перевод символов в формат вероятностных биграмм успешно разрушает устойчивые закономерности языка. Энтропия распределения возрастает, удовлетворяя принципам криптографии Шеннона по максимизации неопределенности. С увеличением числа индексов кривая распределения стремится к «белому шуму» (равномерному распределению), создавая коллизии частот и делая символы неразличимыми для классического анализа. Таким образом, простое увеличение алфавита за счет неизвестного атакующему случайного индексирования существенно осложняет частотный криптоанализ.

ЛИТЕРАТУРА

1. Кромер, В. В. Закон Ципфа и средняя энтропия слова [Электронный ресурс] / В. В. Кромер // Информационные технологии и математическое моделирование : материалы Всерос. науч.-практ. конф. (Анжеро-Судженск, 15 нояб. 2002 г.). – Томск: Твердыня, 2002. – С. 192–194. – Режим доступа: https://www.researchgate.net/publication/336826419_Zakon_Cipfa_i_srednaa_entropia_slova, (дата доступа: 05.02.2026)

2. Бабаш, А. В. Криптография [Электронный ресурс]: учеб. пособие / А. В. Бабаш, Г. П. Шанкин. – Москва : Солон-Р, 2002. – С.

154 - 157. – Режим доступа: https://djvu.online/file/WHhaw5xgDMY9j, (дата доступа: 05.02.2026)

3. А. П. Алферов, А. Ю. Зубов, А. С. Кузьмин, А. В. Черемушкин. Основы криптографии: Учебное пособие. – Гелиос АРВ, 2002. – 480 с.

4. В.А. Долгов, В.В. Анисимов. Криптографические методы защиты информации. – ДВГУПС, 2008. – 155 с.

УДК 004.896 : 658.5

С.А. Осоко, ст. преп;
Е.С. Мирончик, доц., канд. техн. наук (БГТУ, г. Минск)

ТЕХНИЧЕСКИЕ И ЭКОНОМИЧЕСКИЕ ВЫЗОВЫ ПРИ ВНЕДРЕНИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ПРОИЗВОДСТВО

В современном мире искусственный интеллект (ИИ) становится неотъемлемой частью производственных процессов. Однако внедрение ИИ в производство сопряжено с рядом технических и экономических вызовов, которые необходимо учитывать при разработке и реализации соответствующих проектов.

Технические вызовы

1. **Интеграция ИИ с существующими системами**. Многие производственные предприятия используют устаревшие системы и оборудование, которые могут быть несовместимы с новыми технологиями ИИ. Это требует значительных усилий и ресурсов для модернизации и интеграции.

2. **Обеспечение качества и надёжности**. ИИ-системы должны обеспечивать высокое качество и надёжность работы, особенно в критически важных отраслях, таких как производство медицинских препаратов или авиационная промышленность. Необходимо проводить тщательное тестирование и сертификацию ИИ-систем перед их внедрением.

3. **Управление данными**. ИИ требует больших объёмов данных для обучения и работы. Предприятиям необходимо разработать системы сбора, хранения и обработки данных, которые соответствуют требованиям ИИ-систем. Это включает в себя обеспечение качества данных, их защиту и соблюдение нормативных требований.

4. **Разработка и обучение моделей**. Создание эффективных ИИ-моделей требует глубоких знаний и опыта в области машинного обучения, обработки естественного языка, компьютерного зрения и других областях. Предприятиям может потребоваться привлечение