

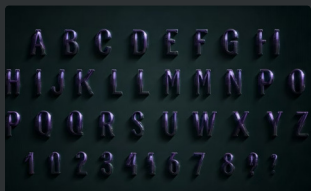


Комбінаторика та ймовірність: розрахунок «часу життя» слабкого пароля

Сценарій заходу: «Вища математика в криптографії та безпеці». Аудиторія — матечно підготовлені слухачі. Мета презентації — пов'язати комбінаторні моделі, ймовірнісні оцінки та поняття ентропії з практикою оцінки міцності паролів та прогнозування часу до зламу (time-to-crack). Структура: вступ (комбінаторика) → комбінаторні моделі стійкості → ймовірнісні інтерпретації і ентропія → практичні застосування → висновки та Q&A.

Вступ: комбінаторика як основа криптографії (3 хвилини)

Паролі розглядаються як елементи дискретних множин: кожен пароль — це k -елементний рядок символів з алфавіту розміром m . Базова теза: стійкість пароля визначається кардинальністю простору паролів. Для моделі розміщень з повтореннями (ordered, repetitions allowed) кількість можливих паролів довжини n дорівнює m^n . Якщо символи мають додаткові обмеження (наприклад, обов'язкові категорії), простір змінюється відповідно. Коротка формула: простір $S = \{\text{all strings of length } n \text{ over alphabet of size } m\}$, $|S| = m^n$.



Алфавіт і довжина

Типовий практичний вибір: $m \approx 26$
(lowercase) + 26 (uppercase) + 10 (digits) + 32
(symbols) $\Rightarrow m \approx 94$. Для $n=8$ маємо $|S| = 94^8$
 $\approx 6.1 \cdot 10^{15}$ можливих рядків.



Обмеження та їхній вплив

Обов'язкова наявність певних класів символів змінює підрахунок: використовується включно-виключно або добуток підпросторів залежно від умов. Наприклад, кількість рядків довжини n , що містять принаймні один символ із заданої підмножини — комбінаторна задача з принципом включення-виключення.

Теза: паролі як елементи дискретних множин — фундаментальні формули

Основні комбінаторні формули (короткі визначення та застосування до паролів): - Розміщення з повтореннями (ordered, repetitions): $|S| = m^n$. - Розміщення без повторень (перестановки довжин n з m символів): $P(m,n) = m!/(m-n)!$ — корисно для політик, що забороняють повтори. - Перестановки всіх символів (повна унікальність): $n!$ для випадку, коли алфавіт рівний довжині та кожен символ використовується рівно один раз. Приклад: якщо $m=94$, $n=8$, то $P(94,8)=94 \cdot 93 \cdot \dots \cdot 87$ — показує значне зменшення порівняно з 94^8 , якщо повтори заборонені.

Важливість моделі

Правильний вибір комбінаторної моделі визначає оцінку складності атак; помилка в моделі (наприклад, припущення про рівномірність) призводить до суттєвого переоцінювання або недооцінювання часу життя пароля.

Приклад числовий

$m=94$, $n=8 \Rightarrow m^n \approx 6.1 \cdot 10^{15}$; при швидкості 10^9 спроб/с (GPU farm) середній час до зламу $\approx (m^n)/2 / \text{rate} \approx 3\,000$ секунд ≈ 50 хвилин — ілюструє ризик слабких політик.

Криптографічна інтерпретація: графи словникових атак

Модель: словникова атака розглядається як обхід графа слів $G=(V,E)$, де вершини — потенційні паролі (або корені словникових варіантів), ребра — операції трансформації (переставлення символів, заміни, додавання цифр). Припустимо, що атака пріоритетизує вершини за ймовірністю p_i (частота використання пароля). Тоді очікуваний час до знаходження цільового пароля залежить від упорядкування вершин за спаданням p_i . Якщо словник охоплює більшу частку маси розподілу (сконцентрована ймовірність), ефективна кардинальність простору значно менша, ніж m^n .



Моделювання пріоритетів

Нехай вершини сортовані так, що $p_1 \geq p_2 \geq \dots \geq p_k$. Якщо атакувальник перевіряє в порядку спадання, то ймовірність зламу до k -ої спроби $= \sum_{i=1}^k p_i$. Отже, необхідна кількість перевірок для досягнення ймовірності встановленої успішності α — мінімальне k з $\sum_{i=1}^k p_i \geq \alpha$.



Фрагментація простору

Розподіл паролів у реальному світі тяжіє до важких хвостів: невелика частка паролів має високу ймовірність (наприклад, "123456", "password"). Це означає, що ефективний простір $S_{\text{eff}} \ll m^n$.

Ентропія Шеннона — математична інтерпретація "часу життя"

Ентропія Шеннона $H(X) = -\sum_i p_i \log_2 p_i$ вимірює середню кількість біт інформації на випадковий вибір пароля за розподілом p . Для рівномірного розподілу $p_i = 1/|S|$ отримуємо $H = \log_2 |S| = n \cdot \log_2 m$. Ця величина еквівалентна кількості незалежних бітів непередбачуваності. Інтерпретація часу життя: якщо середній час перевірки однієї комбінації — τ (секунд), то середній час до зламу в режимі повного перебору при рівномірності $\approx (2^H)/2 \cdot \tau = 2^{H-1} \cdot \tau$.

$$H = n \cdot \log_2(m)$$

Рівномірний випадок

Ентропія прямо пропорційна довжині та логарифму розміру алфавіту.

$$H < n \cdot \log_2(m)$$

Нерівномірний випадок

Популярні паролі зменшують H — ефективна кількість бітів падає, скорочуючи час життя.

Приклад: $m=94, n=8 \Rightarrow H \approx 8 \cdot \log_2(94) \approx 8 \cdot 6.55 \approx 52.4$ біт. Якщо реальний розподіл знижує H до 30 біт, то час життя зменшується в 2^{22} разів ($\sim 4.2 \cdot 10^6$).

Байєсівські оцінки ймовірності зламу

Байєсова перспектива дозволяє інтегрувати апіорні знання про частоту слабких паролів у модель ризику. Формула: $P(\text{злам} | \text{атака}) = [P(\text{атака} | \text{злам}) \cdot P(\text{злам})] / P(\text{атака})$. Інтерпретація: $P(\text{злам})$ — апіорна ймовірність, що обрана жертвою політика породжує слабкий пароль; $P(\text{атака} | \text{злам})$ — ймовірність що атакувальник застосує конкретну стратегію, яка з більшою вірогідністю зламає саме слабкі паролі. Байєсова оновлена ймовірність дозволяє оцінити доцільність захисних заходів (наприклад, блокування IP, примусова зміна пароля).

Орієнтовний числовий приклад

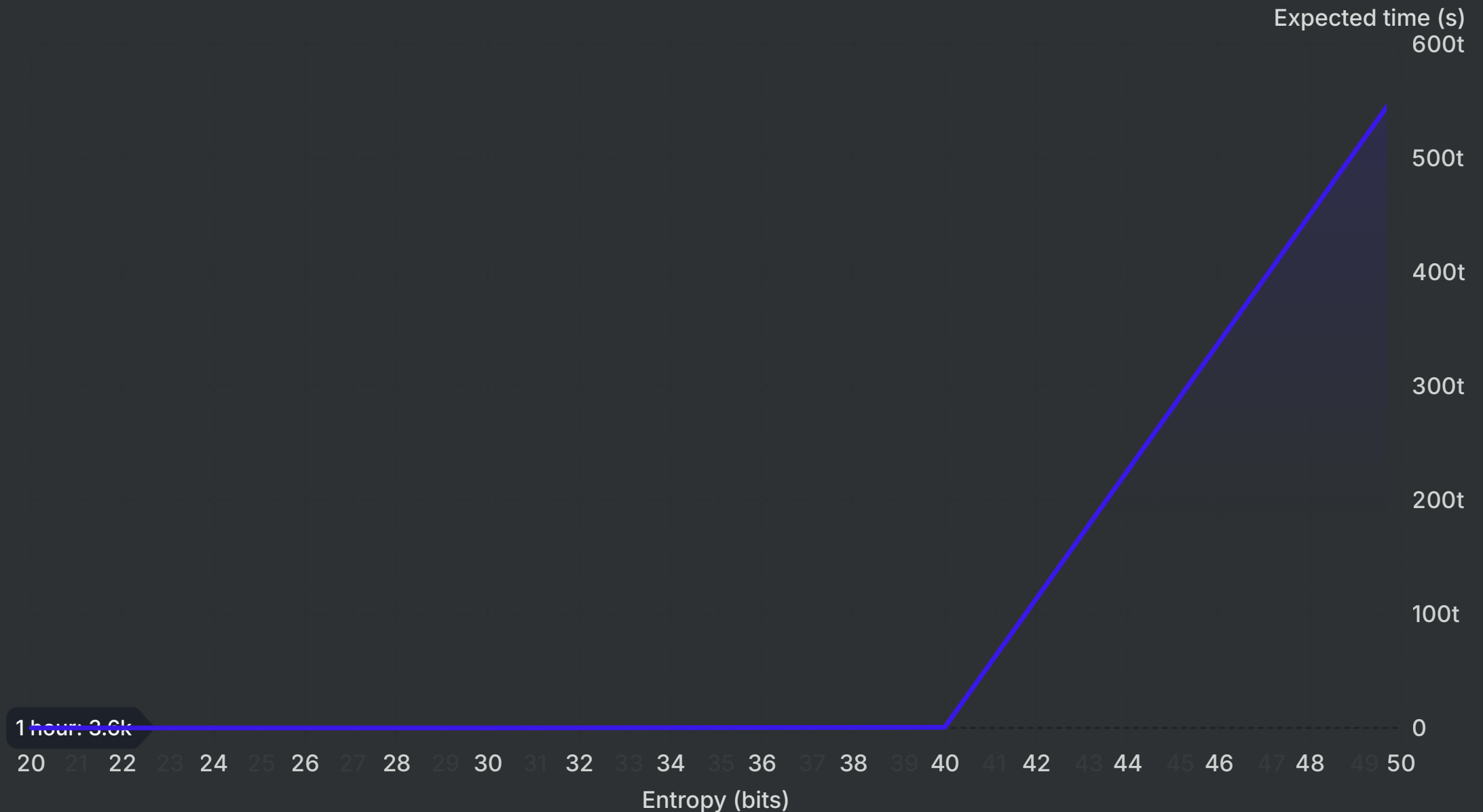
Нехай $P(\text{злам})=0.01$ (1% користувачів обирають слабкий пароль), $P(\text{атака})=0.001$ (ймовірність потрапляння під атаку), $P(\text{атака}|\text{злам})=0.9$. Тоді $P(\text{злам}|\text{атака})= (0.9 \cdot 0.01)/0.001 = 9 \rightarrow$ інтерпретація: за умови атаки ймовірність що вона виявиться успішною дуже висока; насправді результат інтерпретується як перевага умовних подій і вказує на потребу детальнішої нормалізації $P(\text{атака})$.

Практичне застосування

Байєсова модель допомагає пріоритизувати аудит паролів: користувачі з вищою $P(\text{злам})$ підлягають більш суворим вимогам та додатковій автентифікації.

Математична модель «часу життя» (Expected Time To Crack)

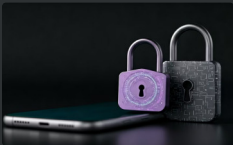
Базова модель повного перебору (brute-force) дає оцінку середнього часу: $E[T] \approx (|S| / 2) / R$, де $|S|$ — розмір простору, R — швидкість перевірок в одиницях спроб/сек. Якщо розподіл паролів не рівномірний, точніша модель використовує геометричний розподіл: Якщо атака має ймовірність вгадати цільовий пароль у кожній незалежній спробі p (рівна для всіх спроб у незмішаній моделі), то час до успіху $T \sim \text{Geometric}(p)$ з $E[T] = 1/p$. Відношення $p \leftrightarrow |S_{\text{eff}}|$: для рівномірного простору $p = 1/|S|$; для концентрованого розподілу p більший для популярних паролів.



Пояснення графіка: експоненційна залежність часу від бітів ентропії. Для даної R значення $E[T] = 2^{N-1}/R$. Невелике падіння N швидко зменшує $E[T]$.

Практичні заходи: як математика покращує безпеку паролів

1) Генератори випадкових чисел: криптографічні PRNG/HRNG забезпечують максимально наближену до рівномірного p_i , збільшують H .
2) Двофакторна автентифікація: комбінаторний добуток просторів — підвищує ефективні біти від H_{password} до $H_{\text{password}} + H_{\text{second_factor}}$ (якщо незалежні). 3) Політики складності: збільшують m та/або n , але також можуть створити людські компроміси (запам'ятовуваність \rightarrow зниження H через предсказуваність). 4) Оцінка вразливостей через умовну ентропію $H(X|Y)$: якщо Y — фрагменти метаданих (наприклад, підказка або персональні дані), умовна ентропія зменшується.



Двофакторна автентифікація

Якщо MFA додає k біт беконфіденційності, загальна ентропія зростає адитивно: $H_{\text{total}} \approx H_{\text{password}} + k$ (при незалежності).



Політика паролів: компроміс

Нав'язування надто складних правил може збільшити кількість повторів (коли користувачі використовують патерни), що фактично зменшує H_{eff} . Рекомендація: довжина $>$ складність.

Алгоритми оцінки ентропії та виявлення слабких паролів

Практичні алгоритми для автоматичної оцінки H_{eff} : - Моделі частоти: навчання на корпусах паролів (p_i оцінюється частотами) → обчислення вибіркової ентропії. - Markov-ланцюги n -го порядку для оцінки лог-імовірності конкретного рядка; швидко і коректно фіксують локальні кореляції символів. - Моделі на основі мовних моделей (NN): оцінюють ймовірність пароля як послідовності символів, дають $\hat{p}(x) \rightarrow \hat{H} = -\log_2 \hat{p}(x)$. - Усереднення по розподілу користувачів дає оцінку середнього H . Алгоритм дії: 1) отримати \hat{p} для кожного кандидата 2) обчислити \hat{H} 3) порівняти з порогом (наприклад, 40 біт) → маркувати як слабкий.

Оцінка через корпуси

Використання відомих витоків (leaked password lists) для отримання апіорних p_i ; швидко і практична, але залежна від релевантності корпусу.

Markov- та n -грами

Підходять для виявлення шаблонів (наприклад, "Year"+"!" або "Qwerty" варіації); забезпечують добрий компроміс між точністю та швидкістю.

Нейромережі

Більш чутливі до складних кореляцій, але потребують ресурси для навчання; дають найкращі оцінки $\hat{p}(x)$ у складних розподілах.

Висновки та Q&A (2 хвилини)

Головна думка: паролі трансформуються з людських рядків в математичні об'єкти; комбінаторика дає кардинальність простору, ентропія вимірює непередбачуваність, а ймовірнісні моделі (включно з байєсовими підходами) дозволяють оцінити реальний ризик та пріоритезувати захисні заходи. Практичні висновки: - Орієнтуйтеся на підвищення ентропії: довші випадкові паролі краще за складні, але короткі. - Використовуйте незалежні фактори автентифікації (MFA) — біти ентропії додаються. - Оцінка H_{eff} повинна базуватись на емпіричних розподілах, а не на теоретичних m^n . - Байєсівський підхід допомагає інтегрувати корисну доменну інформацію в оцінку ризику.



Комбінатори
ка

Ентропія

Практика

Запитання для обговорення: - Як оцінити незалежність між факторами в MFA (адитивність бітів ентропії)? - Які критерії обрати для порога H у різних доменах (банківські сервіси vs. форуми)? - Як коригувати моделі при зміні швидкості атак (незалежно від R)?