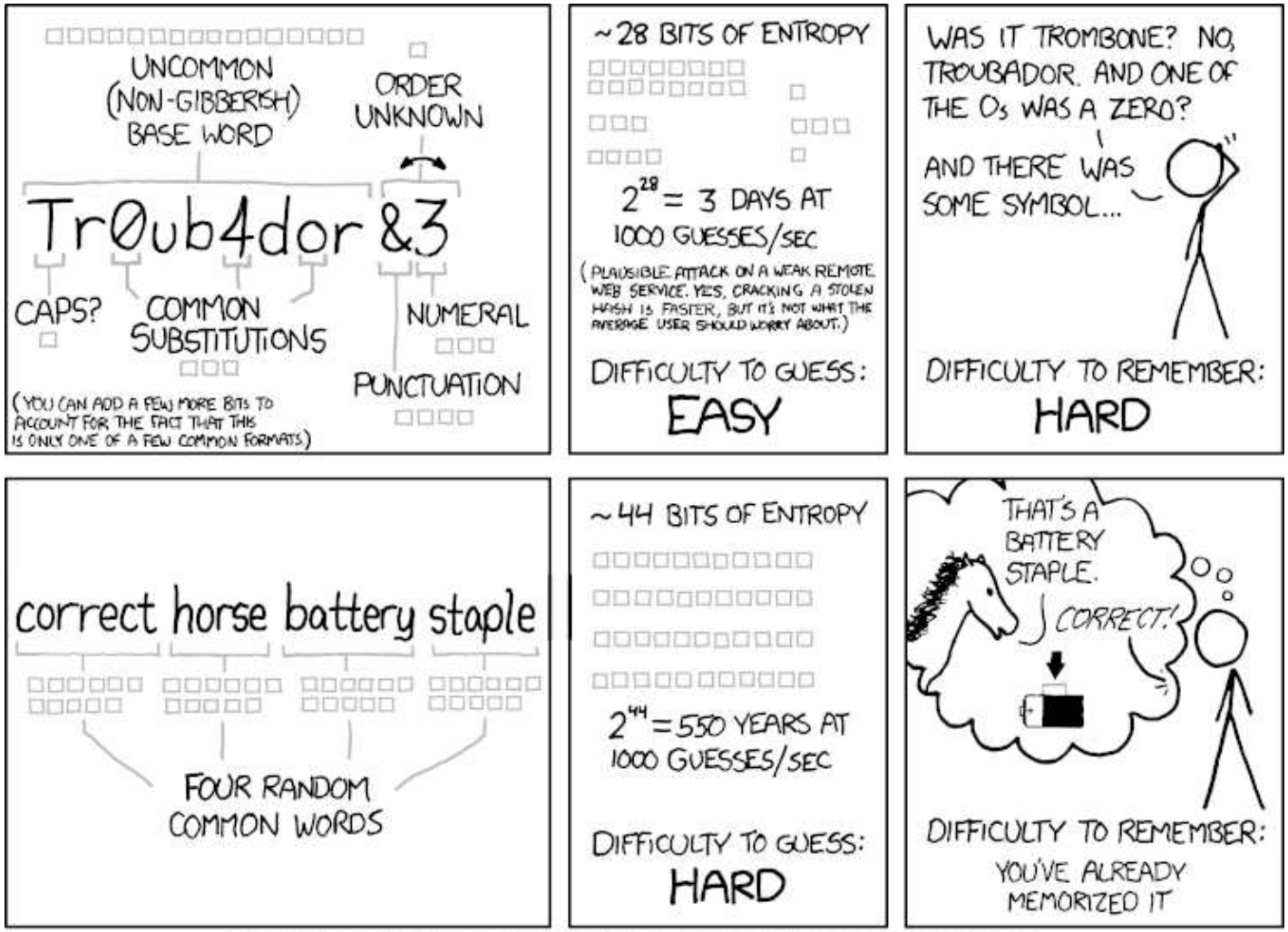
3: Passwörter und Entropie
3.1: Passwörter

- Passwörter (oder Passphrases oder PINs) braucht man dauernd:

- Sie können auch Teil einer "Two-Factor-Authentication" sein (z.B. Chipkarte + Pin).
- Wie wahrscheinlich ist es, dass ein Angreifer mit "wenigen" Versuchen ein Passwort errät?
- **Und wie wählt man ein Passwort, um Angriffe möglichst schwierig zu machen?**

Passwörter im Comic <http://xkcd.com/936/>



Wie viele Versuche hat ein Angreifer denn?

On- und Offline Angreifer

Wir unterscheiden zwischen Online- und Offline-Angreifern

- *Online-Angreifer* (Passworthash $h = H(\mathbf{pw})$ unbekannt):
  - Benötigt Interaktion mit einem Webserver oder Gerät
  - Angreifer an Server: $\mathbf{pw}$
  - Server an Angreifer: *akzeptiert* oder *verworfen*
  - Potentiell hohe Wartezeit nach einigen Versuchen
  - Nur Server/Gerät kann Passworthash vergleichen
  - Nur so viele Versuche wie Server/Gerät zulässt
- *Offline-Angreifer* (Passworthash $h = H(\mathbf{pw})$ bekannt):
  - Kann für beliebige pw' den Hashwert $h' = H(pw')$ berechnen
  - Kann testen ob $h' = h = H(\mathbf{pw})$ gilt
  - Hat beliebig viele Versuche (wird nicht gedrosselt)
  - Fnuktion F: $F(\mathbf{pw}) = 1 \Leftrightarrow \mathbf{pw}$ *akzeptiert*

Offline-Angriffe – gibt es die überhaupt?

1. $\mathbf{pw}$ wird genutzt, um einen geheimen Schlüssel zu berechnen (z.B. für TrueCrypt, um die Daten auf Ihrer Festplatte zu verschlüsseln).
2. Hash($\mathbf{pw}$) wird in einer Benutzer-Datei abgespeichert und dann komprommittiert ... aber das kommt in der Praxis nicht vor! Oder doch? Man google, z.B., nach "passwords compromised":

An Update on LinkedIn Member Passwords Compromised
blog.linkedin.com/2012/.../linkedin-member-**passwords-compromise**...
Jun 6, 2012 – We can confirm that some of the **passwords** that were **compromised** correspond to LinkedIn accounts. We are continuing to investigate this ...

eHarmony Hacked - 1.5 Million Passwords Compromised - Techn...
technorati.com › Technology › Articles
Jun 7, 2012 – Dating website eHarmony joins LinkedIn in suffering from a hack attack, with 1.5 million **passwords** cracked.

Yahoo hacked, 450,000 passwords posted online - CNN.com
www.cnn.com/2012/07/12/tech/web/...**hacked**/index.html

by Doug Gross - in 720 Google+ circles - More by Doug Gross
Jul 13, 2012 – Hackers posted online what they say is login information for more than 450000 Yahoo users.

How to Check if Your Yahoo, Gmail or AOL Passwords Were Leaked
mashable.com/2012/07/12/yahoo-voices-**hacked**/

by Samantha Murphy - in 266 Google+ circles - More by Samantha Murphy
Jul 12, 2012 – Yahoo has been the subject of a massive data breach, and your email address may be among the thousands **compromised**.

Blizzard's Battle.net Hacked; Company Recommends All Users ...
www.macrumors.com/.../blizzards-battle-net-**hacked**-company-recom...
Aug 9, 2012 – Blizzard Entertainment, the company behind Warcraft, Starcraft and Diablo, today informed customers that their internal security network had ...

Searches related to **passwords compromised**
passwords compromised **yahoo**
apache project server hacked passwords compromised
msn email compromised
was my linkedin password compromised

Gibt es Alternativen zu Passwörtern?

Drei mögliche Faktoren, die Identität einer Person festzustellen

Wissen: Passwörter, PINs, ...

Besitz: Smartcards, Dongles, ...

- Teuer
- Diebe in der physikalischen Welt
- Unpraktisch

Eigenschaft: Biometrische Merkmale, z.B. Fingerabdrücke, Retina-Scans, Bewegungsmuster, ...

- Unzuverlässig ("false rejectance")
- Daten (z.B. Fingerabdrücke) sind nicht geheim
- Weitgehend nutzlos für Online-Authentifikation

Two-Factor-Authentiation

- Bei höhere Sicherheitsanforderungen **verbindet** man zwei der obigen Ansätze.
- Meistens Wissen und Besitz.
- Manchmal Wissen und Eigenschaften.
- Lösungen ohne den Faktor Wissen sind selten.

Passwort-Scrambling

Ist noch was zu retten?

- *früher:* (username, **pw**, . . .) in Datei
- *UNIX* `crypt` (username, S, H(S, **pw**), . . .)
  - 25 Wiederholung einer DES-artigen Operation (“Key Stretching”)
  - 12-Bit Salt S (gegen Wörterbuch-Angriffe)
- *1988:* “shadow passwords”: separate Datei mit H(**pw**, S)
- *1995:* Abadi, Lomas, Needham: “pepper”
  (einige Bits von S bleiben unbekannt)
- *1997:* Kelsey, Schneier, Hall, Wagner: i Iterationen (“key strechting”)
- *2007:* Boyen: Unbekannte Anzahl an Iterationen (“halting”)
- *2010:* Percival: Erste speicherintensive KDF **`scrypt`**
- *2013:* Aufruf zu einem internationalen Wettbewerb (PHC) für moderne Password Scrambler
- *2015:* Argon2 gewinnt PHC

Einige Passwort-Regeln und -Mythen

Keine Wörter aus dem Wörterbuch: Gute Regel!

Groß- und Kleinschreibung, Ziffern und Sonderzeichen:
Schadet nichts – wenn man sich das korrekte Passwort trotzdem noch merken kann. Wird in vielen Systemen beim Passwort-Wechsel erzwungen.
Ist aber bei hinreichend langen Passwörtern überflüssig!

W0rt3r verfr3mden: Schadet nichts, bringt auch nicht viel.
Den Trick kennen Passwort-Cracker schon lange!

Passwörter nicht aufscheiben: Besser: Passwort aufschreiben und gut verstecken! Dann traut man sich eher, tatsächlich ein starkes Passwort zu wählen. Hat man das Passwort oft genug verwendet um es sich eingeprägt zu haben, sollte man den Zettel vernichten.

Welche dieser Passwörter sind gute Passwörter?

- “123456” (das häufigste Passwort überhaupt)?
- “the magic words are squeamisch ossifrage”?
- “squeamish:ossifrage”?
- “SusiamStrandvonSassnitzimJuli2007”?
- “Yxcvbnmasdfghjklqwertzuiop2$”?
- “2a94m:-5%ruxkelsllah65!”?

Keines der genannten Passwörter ist stark!

- Einige sind offensichtlich schwach.
- Einige sind vielleicht **stark gewesen**, bevor sie auf meinen Folien veröffentlicht wurden – aber jetzt sind sie öffentlich bekannt und damit schwach!

- Wir wollen nach einem **systematischen Verfahren** suchen, **gute zufällige Passwörter** zu erzeugen.
- Wir wollen die Qualität des Verfahrens quantifizieren – und die tatsächlich gewählten Passwörter geheim halten.
- Wichtig ist die **Wahrscheinlichkeit**, ein zufällig gewähltes Passwort zu erraten.

Wie gut sollte ein ~~Passwort~~ Verfahren sein?

Faustregeln

$\Pr[\mathbf{pw}]$	sicher gegen	Sicherheit
$\approx 2^{-10}$	Online-Angreifer	schwach
$\approx 2^{-30}$		stark
$\approx 2^{-50}$	Offline-Angreifer	schwach
$\approx 2^{-70}$		stark
$\approx 2^{-90}$		sehr stark

- "Schwache" Sicherheit: Basis-Schutz vor Gelegenheitsangriffen
- "Starke" Sicherheit: Schutz vor den meisten professionellen Angriffen (typische Bot-Netze, ...)
- "Sehr starke" Sicherheit: Schutz vor allem, was heute vorstellbar ist (Geheimdienste, ...)

Gleichverteilt und zufällig

Wähle ein zufälliges Passwort als String einer vorgegebenen Länge aus einem der folgenden "Alphabete"

1. Dezimalziffern (10): "37823431"
2. Kleinbuchstaben "jewkjymh"
3. Die "druckbaren" ASCII-Zeichen (95): "ad%f9we:"
4. Wörter aus einem Wörterbuch (z.B. 38 000): "celebration stopper nest ailment chimneysweep specification toneless deviate"

Konkrete Rechnung

Wieviele Zeichen braucht ein zufälliges Passwort über einem Alphabet der Größe 10, 26, 95, bzw. 38 000, um eine vorgegebene Wahrscheinlichkeit zu unterschreiten?

$\Pr[\mathbf{pw}]$	1.	2.	3.	4.
α	10	26	95	38000
$< 2^{-10}$	4	3	2	1
$< 2^{-30}$	10	7	5	2
$< 2^{-50}$	16	11	8	4
$< 2^{-70}$	22	15	11	5
$< 2^{-90}$	29	20	14	6

Ziemlich lang!

Wie sollen sich Nutzer **zufällige** Passwörter mit einer "vernünftigen" Sicherheit gegen Offline-Angriffe überhaupt merken?

Die Empfehlung des BSI

Bundesamt für Sicherheit in der Informationstechnik

"Weil so ein Passwort aber schwer zu merken ist, rät das BSI zu einem Merksatz als Eselsbrücke. Aus "Ich habe 100 sichere Passwörter, um mich im Internet anzumelden!" wird etwa "Ih100sP,umiIa!". Der Merksatz muss unbedingt selbst ausgedacht sein. Wer Anfangsbuchstaben von bekannten Phrasen, Sätzen, Liedtexten oder Gedichten wählt, läuft Gefahr, dass Angreifer diese bereits in ihren Wörterbuch-Attacken berücksichtigen."

– Die Welt, 04. Feb. 2013

Was ist von dieser Empfehlung zu halten?

3.2: Entropie

Die "**Entropie**", soll den *Informationsgehalt* einer Nachrichtenquelle angeben.

> **Entropie:** Anzahl Bits die mindestens gebraucht werden, um die Information einer Nachrichtenquelle darzustellen. (Intuition, noch keine Definition!)

Kurz: Entropie = Mindest-Anzahl Bits zur Speicherung

- Eine Quelle, die permanent immer die gleiche Information versendet, verbreitet keine Information.
- Eine Quelle, die das Ergebnis "Kopf" oder "Zahl" eines fairen Münzwurfs liefert, hat 1 bit Entropie.

Was hat die Länge der (komprimierten) Darstellung mit der Sicherheit eines Passwortes zu tun?

Wenn man zum Speichern eines unbekannten Wertes mindestens b bit brauche, und ich versuche, diesen Wert zu erraten, dann kann ich statistisch bestenfalls erwarten, dies nach 2^{b-1} Versuchen zu schaffen.

Die Entropie einer Nachrichtenquelle (für uns insbesondere eines Verfahrens, Passwörter zu generieren!) gibt an, welchen Aufwand man *im Durchschnitt* treiben muss, um ein Passwort aus dieser Quelle zu "knacken".

Beispiel

Wie viele Bits braucht man mindestens, um die Wörter der folgenden Quelle zu speichern?

Die Quelle erzeugt die Buchstaben “a”, bis “d”, und zwar

- “a” mit der Wahrscheinlichkeit $\Pr[\text{“a”}] = 0.5$,
- “b” mit der Wahrscheinlichkeit $\Pr[\text{“b”}] = 0.25$, und
- “c” und “d” mit $\Pr[\text{“c”}] = \Pr[\text{“d”}] = 0.125$?

Variante I

a	$\leftrightarrow$	00	b	$\leftrightarrow$	01
c	$\leftrightarrow$	10	d	$\leftrightarrow$	11

1 Buchst. $\leftrightarrow$ 2 bit.

Wir brauchen 2 bit pro Zeichen, für 100 Zeichen also 200 bit.

Variante II

a	$\leftrightarrow$	1	b	$\leftrightarrow$	01
c	$\leftrightarrow$	000	d	$\leftrightarrow$	001

1 Buchst. $\leftrightarrow$ 1.75 bit.

Die Entropie dieser Quelle beträgt also höchstens 1.75 bit.

Anders ausgedrückt: Um eine Folge von 100 Zeichen darzustellen, werden im Durchschnitt 175 bit gebraucht.

Ist die Entropie der Quelle vielleicht noch kleiner, d.h., können wir 100 Zeichen vielleicht noch kompakter darstellen?

Shannon-Entropie

Definition 7 (Shannon-Entropie)

Sei Q eine Informationsquelle mit $|Q| = n$ und seien $p_i \in Q, 1 \leq i \leq n$ die Wahrscheinlichkeiten des Auftretens der Einzelereignisse in Q. Dann ist die Shannon-Entropie einer Quelle Q gegeben durch:

$$H_1(Q) = -\sum_i p_i \cdot \log_2(p_i).$$

- Die Shannon-Entropie gibt den mittleren Informationsgehalt einer Quelle Q an.
- **Fairer Würfel:** Es gilt: $p_i = 1/6$ und $\log_2(1/6) \approx -2.58$.

$$H_1(\textbf{Fairer Würfel}) = -(6 \cdot (1/6 \cdot (-2.58)) = \log_2(6) = \mathbf{2.58} \text{ bit}$$

Entropie bei Gleichverteilung

Satz 8

Ist Q eine gleichverteilte Quelle von n Elementen, dann ist $H_1(Q) = \log_2(n)$.

Beweis.

Ist Q gleichverteilt, also $\mathsf{Pr}[q_i] = 1/n$, dann ist

$$\begin{aligned} H_1(Q) &= -\sum_{i=1}^{n} \mathsf{Pr}[q_i] * log_2(\mathsf{Pr}[q_i]) = -\sum_{i=1}^{n} \frac{1}{n} \log_2(1/n) \\ &= -n * \frac{1}{n} * \log_2 \frac{1}{n} = -\log_2\left(\frac{1}{n}\right) = \log_2(n) \end{aligned}$$

□

Entropie bei Ungleichverteilung

Satz 9

Ist Q nicht gleichverteilt, so ist

$$H_1(Q) < \log_2(n).$$

(ohne Beweis)

Folgerung

Egal wie Q verteilt ist, stets gilt

$$0 \leq H_1(Q) \leq \log_2(n).$$

Entropie bei unabhängigen Zufallsquellen

Für zwei Quellen Q und R bezeichnet QR die Menge zufälliger Paare (q, r), von Elementen q aus Q und r aus R.

Satz 10

Sind Q und R zwei voneinander unabhängige Quellen, dann ist

$$H_1(QR) = H_1(Q) + H_1(R).$$

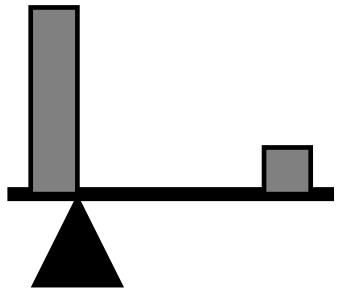
Es bezeichne Q^i die Quelle, die durch i-faches und unabhängiges Ziehen eines Elements aus Q definiert ist.

Folgerung

$$H_1(Q^i) = i * H_1(Q).$$

Die Min-Entropie

- Die Shannon-Entropie fragt nach der *durchschnittlichen* Anzahl an Bits, die man braucht, um die Wörter aus einer Quelle zu speichern.
- Eine Eigenschaft des *Durchschnitts* ist jedoch, dass es Konstellationen gibt, bei denen fast alle Werte unterdurchschnittlich und wenige überdurchschnittlich sind (siehe Abb. rechts).



- Bezogen auf die Passwort-Sicherheit kann das zu optimistischen Schlussfolgerungen führen.
- Deshalb betrachtet man alternativ zur Shannon-Entropie auch die Min-Entropie.

Die Min-Entropie

als "pessimistische" Alternative zur Shannon-Entropie

Sei Q eine Quelle ...

Definition 11 (Min-Entropie)

Die **Min-Entropie** von Q ist

$$H_\infty(Q) = \min_i(-\log_2(\Pr[q_i])) = -\log_2(\max(\Pr[q_i])).$$

Satz 12

Ist Q gleichverteilt, gelten

$$H_\infty(Q) = H_1(Q) = \log_2(n) \quad \textit{und} \quad 0 \le H_\infty(Q) \le H_1(Q) \le \log_2(n),$$

außerdem

$$H_\infty(Q^i) = i \cdot H_\infty(Q).$$

3.3: Merksatz-Passworte und ihre Entropie

- Zur Erinnerung: Wir verwenden einen Merksatz wie “Ich habe 100 sichere Passwörter, um mich im Internet anzumelden!”
- Wir versuchen, die Entropie der daraus abgeleiteten Passwörter abzuzchätzen.
- Wir vereinfachen das Modell so weit, dass wir plausibel (?) annehmen können, dass die einzelnen Buchstaben weitgehend **unabhängig** voneinander sind. Nur dann können wir aus der Entropie eines Buchstabens auf die Entropie des Passwortes insgesamt schließen (egal für welche Art der Entropie):

$$H(Q^i) = i \cdot H(Q).$$

- Satzzeichen, Ziffern, und der Wechsel von Groß- und Kleinschreibung hängen alle voneinander ab, deshalb schränken wir uns auf **Kleinbuchstaben** ein; unser aus dem obigen Satz abgeleitetes Passwort wird dann zu “ihhspumiia”.

- 26 Buchstaben ‘a’, . . . , ‘z’
- Bei Gleichverteilung: Entropie $\log_2(26) \approx 4.7$ bit pro Buchstabe.
- Empirische Berechnung der Entropie:
  - viele Texte von vielen verschiedenen Autoren;
  - zähle die Häufigkeit verschiedener Anfangsbuchstaben (und ignoriere Umlaute),
  - berechne die empirischen Wahrscheinlichkeiten der Buchstaben,
  - berechne Shannon- und Min-Entropie H_1 und H_∞:

$$H_1(Q) = -\sum_{i=1}^{n} \Pr[q_i] \log_2 \Pr[q_i]$$

$$H_\infty(Q) = \min_i(-\log_2(\Pr[q_i])) = -\log_2(\max(\Pr[q_i])).$$

  - Bei der Shannon-Entropie ist zu beachten: $\lim_{\Pr[q_i]\to 0} \Pr[q_i] \cdot \log_2(\Pr[q_i]) = 0$, deshalb ist für $\Pr[q_i] = 0$ definiert: $\Pr[q_i] \cdot \log_2(\Pr[q_i]) = 0$.
- Welche Shannon- und Min-Entropie erhalten wir so?

Das tatsächliche Experiment

Nicht sehr wissenschaftlich: Ein einziger Text von einem einzigen Autor

Friedrich Schiller

Was heißt und zu welchem Ende studiert man Universalgeschichte?

Eine akademische Antrittsrede

Erfreuend und ehrenvoll ist mir der Auftrag, meine h. H. H., an Ihrer Seite künftig ein Feld zu durchwandern, das dem denkenden Betrachter so viele Gegenstände des Unterrichts, dem tätigen Weltmann so herrliche Muster zur Nachahmung, dem Philosophen so wichtige Aufschlüsse und jedem ohne Unterschied so reiche Quellen des edelsten Vergnügens eröffnet – das große weite Feld der allgemeinen Geschichte. Der Anblick so vieler vortrefflichen jungen Männer, die eine edle Wißbegierde um mich her versammelt und in deren Mitte schon manches wirksame Genie für das kommende Zeitalter aufblüht, macht mir meine Pflicht zum Vergnügen,

...

Die Ergebnisse

a:	415	b:	271	c:	68
d:	**966**	e:	536	f:	182
g:	355	h:	251	i:	328
j:	64	k:	144	l:	133
m:	267	n:	242	o:	36
p:	58	q:	5	r:	165
s:	**646**	t:	142	u:	421
v:	310	w:	435	x:	0
y:	0	z:	235		

Dies sind **insgesamt 6675 Wörter**. Der häufigste Anfangsbuchstabe ist das ‘d’, er kommt 966-mal vor. Der zweithäufigste ist das ‘s’.

Shannon-Entropie: $H_1(Q) \approx 4.18$ bit, Min-Entropie: $H_\infty \approx 2.79$ bit.

Kann man das verbessern?

Mit knapp 2.8 bit pro Buchstabe ist die Min-Entropie beunruhigend klein!

Beobachtung:

- Die bestimmten Artikel “der”, “die” und “das” kommen zusammen 402-mal vor.
- Ihr Beitrag zur Entropie ist gering.

Also:

- “denken” wir uns die bestimmten Artikel zu unseren Merksätzen hinzu,
- ignorieren sie aber für die Passwörter (bzw. die Berechnung der Entropie).

Bei gleicher Passwort-Länge

- brauchen wir etwas längere Merksätze,
- bekommen aber stärkere Passwörter.

Ohne “der”, “die” und “das”: Mehr Entropie!

a:	415	b:	271	c:	68
d:	**564**	e:	536	f:	182
g:	355	h:	251	i:	328
j:	64	k:	144	l:	133
m:	267	n:	242	o:	36
p:	58	q:	5	r:	165
s:	**646**	t:	142	u:	421
v:	310	w:	435	x:	0
y:	0	z:	235		

Dies sind **insgesamt 6273 Wörter**. Das ’d’ ist nur noch der zweithäufigste Buchstabe, das ’s’ ist mit 646 Vorkommen nun der häufigste.

Shannon-Entropie: $H_1(Q) \approx 4.25$ bit, Min-Entropie: $H_\infty \approx 3.28$ bit.

Was bedeutet das in der Praxis?

Unter der Annahme, dass die Daten des Schiller-Textes auch für moderne deutsche Texte gelten und die Anfangsbuchstaben in deutschen Sätzen statistisch (annähernd] unabhängig sind

- Das Verfahren

  Merksatz $\rightarrow$ Anfangsbuchstaben ohne “der”, “die” und “das”

  scheint starke Passwörter zu liefern – obwohl diese nur aus Kleinbuchstaben bestehen.

- Wichtig ist, dass das Passwort lang genug ist:

Ziel	**# Buchstaben**	**Gleichverteilt**
$H_\infty(\mathbf{pw}) \geq 90$	28	20
$H_\infty(\mathbf{pw}) \geq 70$	22	15
$H_\infty(\mathbf{pw}) \geq 50$	16	11
$H_\infty(\mathbf{pw}) \geq 30$	10	7

3.4: Schlussbemerkungen

- Nutzern fällt es schwer, gute Passwörter zu wählen.
- Die üblichen Regeln (mindestens X Zeichen, Groß - und Kleinbuchstaben, Ziffern und Satzzeichen mischen, ...) führen nicht immer zu starken Passwörtern – dafür oft zu solchen, die schwierig sind, sich zu merken!
- Die Stärke von Passwörtern (genauer, die Stärke von Verfahren, Passwörter zu erzeugen) kann man mit der **Entropie** quantifizieren.
- Sie sollten insbesondere
  - anhand verschiedener Angriffszenarien (Offline vs. Online) erklären können, wann man sehr starke Passwörter braucht, wann ggf. weniger starke,
  - und mit der Schannon- sowie die Min-Entropie von Nachrichtenquellen rechnen bzw. sie berechnen können.