Informationsteori

Daniel Bosk[1]

Avdelningen för informations- och kommunikationssytem (IKS),
Mittuniversitetet, Sundsvall.

infotheory.tex 998 2013-05-01 02:04:04Z danbos

Översikt

1. Informationsteori
   - Historia
   - Definition av Shannonentropi
   - Jensens olikhet
   - Egenskaper för Shannonentropi
   - Betingad entropi
   - Informationstäthet och redundans
   - Informationsvinst
2. Tillämpningar inom säkerhet
   - Lösenord
   - Identifierande information
   - Kryptografi

Litteratur

Innehållet i föreläsningen motsvarar

- några kortare skrifter om entropi [Uel; Eck10b; Eck10a],
- kompendiet [Bos13], samt
- några artiklar om lösenord [KRC06; KSK+11].

Översikt

Mittuniversitetet
MID SWEDEN UNIVERSITY

Översikt

Mittuniversitetet
MID SWEDEN UNIVERSITY

Historia

- Området informationsteori skapades 1948 genom Shannons artikel ”A Mathematical Theory of Communication” [Sha48].
- I sin artikel myntar han begreppet entropi för information.
  - Begreppet entropi fanns sedan tidigare inom fysiken.
  - Myntades där av John von Neumann, som även föreslog namnet för Shannon.
- Han använder sin teori för att analysera kommunikation.
  - Vad är de teoretiska gränserna för kommunikationskanaler?
  - Både med och utan brus, hur mycket redundans krävs för given nivå brus?

Översikt

Definition av Shannonentropi

Definition (Shannonentropi)

Låt $\mathbf{X}$ vara en stokastisk variabel som antar värden från en ändlig mängd X. Då definieras *Shannonentropin* för den stokastiska variabeln $\mathbf{X}$ som

$$H(\mathbf{X}) = -K \sum_{x \in X} \Pr(\mathbf{X} = x) \log \Pr(\mathbf{X} = x),$$

där K oftast väljs som $\frac{1}{\log 2}$. Då sägs entropin anges i enheten bitar (bit).

Definition av Shannonentropi

- Vi kan se Shannonentropin för en stokastisk variabel som hur mycket ”val” som varje utfall innebär.
- Den kan även tolkas som osäkerheten för ett utfall för variabeln.
- Eller hur många bitar som krävs för att lagra resultatet av ett utfall.
- Och då naturligtvis ett mått på hur mycket information den producerar.

Definition av Shannonentropi

Exempel (Singla slant)

Låt **S** beteckna en stokastisk variabel som antar värden ur mängden $S = \{krona, klave\}$. Då har vi att

$$\Pr(\mathbf{S} = krona) = p_{krona} = \Pr(\mathbf{S} = klave) = p_{klave} = \frac{1}{2}.$$

Entropin för **S** är följaktligen

$$\begin{aligned} H(\mathbf{S}) &= -\left(p_{krona} \log p_{krona} + p_{klave} \log p_{klave}\right) \\ &= -2 \times \frac{1}{2} \log \frac{1}{2} = \log 2 = 1. \end{aligned}$$

Definition av Shannonentropi

Exempel (Kasta tärning)

Låt $\mathbf{T}$ beteckna en stokastisk variabel som antar värden ur mängden $T = \{1, 2, 3, 4, 5, 6\}$. Då har vi att $\Pr(\mathbf{T} = t) = \frac{1}{6}$ för alla $t \in T$. Följaktligen är entropin för $\mathbf{T}$

$$H(\mathbf{T}) = -\sum_{t \in T} \Pr(\mathbf{T} = t) \log \Pr(\mathbf{T} = t)$$

$$= -6 \times \frac{1}{6} \log \frac{1}{6} = \log 6 \approx 2.585.$$

Definition av Shannonentropi

- Betyder alltså att det svårare att förutspå ett tärningskast än slantsingling.
- För att lagra resultaten av tärningskasten krävs mer utrymme.
- Låt oss förvanska tärningen en aning . . .

Definition av Shannonentropi

Exempel (Tärningskast igen)

En ny stokastisk variabel T' antar värden ur $T = \{1, 2, 3, 4, 5, 6\}$. Låt nu $\Pr(\mathbf{T}' = 6) = \frac{9}{10}$ och $\Pr(\mathbf{T}' = t) = \frac{1}{5 \times 10}$ för alla $t \neq 6$. Detta ger entropin

$$
\begin{aligned}
H(\mathbf{T}') &= -\left(\frac{9}{10} \log \frac{9}{10} + \sum_{t \neq 6} \frac{1}{50} \log \frac{1}{50} \right) \\
&= -\frac{9}{10} \log \frac{9}{10} - 5 \times \frac{1}{50} \log \frac{1}{50} \\
&= -\frac{9}{10} \log \frac{9}{10} - \frac{1}{10} \log \frac{1}{50} \approx 0.486.
\end{aligned}
$$

Översikt

Jensens olikhet

Definition

Låt $f : \mathbb{R} \to \mathbb{R}$ vara en funktion sådan att

$$tf(x) + (1-t)f(y) \leq f(tx + (1-t)y),$$

då säger vi att f är *konkav*. Vid strikt olikhet för $x \neq y$ säger vi att f är *strikt konkav*.

Exempel

log är en strikt konkav funktion.

Jensens olikhet

Sats (Jensens olikhet)

Om f är en strikt konkav funktion och $a_1, a_2, \ldots, a_n$ är reella tal strikt större än noll och sådana att $\sum_{i=1}^{n} a_i = 1$ då gäller att

$$\sum_{i=1}^{n} a_i f(x_i) \leq f\left(\sum_{i=1}^{n} a_i x_i\right),$$

med likhet om och endast om $x_1 = x_2 = \cdots = x_n$.

Jensens olikhet

Bevis för Jensens olikhet.

Bevis genom induktion. Antag $n = 2$. Då har vi att $a_1 + a_2 = 1$ och alltså $a_1 = 1 - a_2$. Eftersom f är konkav har vi att

$$a_1 f(x_1) + a_2 f(x_2) \leq f(a_1 x_1 + a_2 x_2).$$

Antag sant för $n = k$.

$$\sum_{i=1}^{k} a_i = 1 \wedge \sum_{i=1}^{k} a_i f(x_i) \leq f\left(\sum_{i=1}^{k} a_i x_i\right). \tag{1}$$

Jensens olikhet

Forts. bevis för Jensens olikhet.

Lås oss visa att detta även gäller för $n = k + 1$. Eftersom att f är konkav gäller att

$$\sum_{i=1}^{k+1} a_i f(x_i) = a_1 f(x_1) + (1 - a_1) \sum_{i=2}^{k+1} \frac{a_i}{1 - a_1} f(x_i) \tag{2}$$

$$\leq f\left(a_1 x_1 + (1 - a_i) \sum_{i=2}^{k+1} \frac{a_i}{1 - a_1} x_i \right). \tag{3}$$

Då $\sum_{i=2}^{k+1} \frac{a_i}{1-a_i} = 1$ kan vi tillämpa induktionshypotesen (1) i (2) och (3), alltså är det sant för alla $n \in \mathbb{N}$.
Vidare följer likhet från att f är strikt konkav. Q.E.D.

Översikt

Egenskaper för Shannonentropi

Sats

Låt $\mathbf{X}$ *vara en stokastisk variabel med sannolikhetsdistribution som antar värdena* $p_1, p_2, \ldots, p_n$, *där* $p_i > 0$ *för* $1 \leq i \leq n$. *Då är* $H(\mathbf{X}) \leq \log n$, *med likhet om och endast om* $p_1 = p_2 = \cdots = p_n = 1/n$.

Egenskaper för Shannonentropi

Bevis.

Satsen följer direkt från Jensens olikhet:

$$H(\mathbf{X}) = -\sum_{i=1}^{n} p_i \log p_i = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$$
$$\leq \log \sum_{i=1}^{n} p_i \frac{1}{p_i} = \log n.$$

Med likhet om och endast om $p_1 = p_2 = \cdots = p_n$. Q.E.D.

Egenskaper för Shannonentropi

Korollarium

$H(\mathbf{X}) = 0$ *om och endast om* $\Pr(\mathbf{X} = x) = 1$ *för något* $x \in X$ *och* $\Pr(\mathbf{X} = x') = 0$ *för alla* $x \neq x' \in X$.

Bevis.

Om $\Pr(\mathbf{X} = x) = 1$ då är $n = 1$ och således $H(\mathbf{X}) = \log n = 0$.
Om $H(\mathbf{X}) = 0$, då måste $\log n = 0$. Följaktligen är $n = 1$. Q.E.D.

Egenskaper för Shannonentropi

Sats

Följande egenskaper gäller:

1. *H är kontinuerlig.*
2. *Om* $\Pr(\mathbf{X} = x) = 1/|X|$ *för alla* $x \in X$ *då är H en monotont stigande funktion med avseende på* $|X|$.

Bevis.

(1) följer direkt av att logaritmen är kontinuerlig och funktionssammansättningar av kontinuerliga funktioner är kontinuerliga.
(2) följer av föregående sats. Q.E.D.

Egenskaper för Shannonentropi

Lemma

Låt $\mathbf{X}$ *och* $\mathbf{Y}$ *vara stokastiska variabler. Då gäller att*

$$H(\mathbf{X}, \mathbf{Y}) \leq H(\mathbf{X}) + H(\mathbf{Y}),$$

med likhet om och endast om $\mathbf{X}$ *och* $\mathbf{Y}$ *är oberoende.*

Översikt

Mittuniversitetet
MID SWEDEN UNIVERSITY

Betingad entropi

Definition (Betingad entropi)

Vi definierar den *betingade entropin* $H(\mathbf{Y} \mid \mathbf{X})$ som

$$H(\mathbf{Y} \mid \mathbf{X}) = -\sum_{y}\sum_{x} \Pr(\mathbf{Y} = y) \Pr(\mathbf{X} = x \mid y) \log \Pr(\mathbf{X} = x \mid y).$$

Denna mäter osäkerheten hos $\mathbf{X}$ som inte avslöjas av $\mathbf{Y}$.

Betingad entropi

Sats

$H(\mathbf{X}, \mathbf{Y}) = H(\mathbf{Y}) + H(\mathbf{X} \mid \mathbf{Y})$

Korollarium

$H(\mathbf{X} \mid \mathbf{Y}) \leq H(\mathbf{X})$.

Översikt

Informationstäthet och redundans

Definition

Låt L vara ett naturligt språk och $\mathbf{P}^n$ vara en stokastisk variabel som har sannolikhetsfördelningen av strängar av längd n i språket L. Vi definierar *entropin* för språket L att vara

$$H_L = \lim_{n \to \infty} \frac{H(\mathbf{P}^n)}{n}$$

och *redundansen* för L att vara

$$R_L = 1 - \frac{H_L}{\log |P|}.$$

Informationstäthet och redundans

- Detta betyder att vi har H_L bitar information per tecken i språket L.
- Experiment har visat att entropin för engelska är mellan 1 och 1.5 bitar per tecken.
- Redundansen blir följaktligen ungefär $1 - \frac{1.25}{\log 26} = 0.61$.
- Shannon påpekar i sin artikel [Sha48] att redundansen måste vara omkring 0.5 för att tvådimensionella korsord ska vara meningsfulla.
- Redundansen i "SMS-svenska" är lägre än i vanlig svenska: jämför "också" med "oxå".

Informationstäthet och redundans

- Detta säger också att vi kan uppskatta entropin för en given Markovprocess.
- Shannon modellerade språket som en Markovprocess i sin artikel [Sha48].
- Vi kan även beräkna entropin för ett givet tillstånd i en Markovprocess genom betingad entropi.

Översikt

Mittuniversitetet
MID SWEDEN UNIVERSITY

Informationsvinst

Definition

Låt U vara mängden av möjliga utfall och sannolikheten för utfall i betecknas p_i. Om vi får veta att utfallet är i en delmängd $A \subset U$ definierar vi *informationsvinsten* $G(A \mid U)$ som

$$G(A \mid U) = \log \frac{1}{\Pr(A)} = -\log \Pr(A),$$

där $\Pr(A) = \sum_{i \in A} p_i$.

Informationsvinst

Exempel (Tärningskast igen)

Vi gör ett kast med en perfekt tärning. Om vi får veta att tärningskastet är ett jämnt tal har vi fått

$$-\log\left(\frac{1}{6}+\frac{1}{6}+\frac{1}{6}\right) = -\log\frac{3}{6} = \log\frac{6}{3} = \log 2 = 1$$

bit informaton. Osäkerheten som återstår är således ungefär 1.58 bitar.

Informationsvinst

Exempel (Ytterligare tärningskast)

Vi får veta att tärningen visar mindre än fem. Då har vi fått

$$-\log\left(5 \times \frac{1}{6}\right) = \log\frac{6}{5} \approx 0.26$$

bitar information.

Översikt

Översikt

Mittuniversitetet
MID SWEDEN UNIVERSITY

Lösenord

Idé

Kan använda $H(x_1, x_2, \ldots, x_n) = H(x_1) + H(x_2) + \cdots + H(x_n)$ för att räkna på entropin hos lösenord, där x_i är olika egenskaper hos lösenordet.

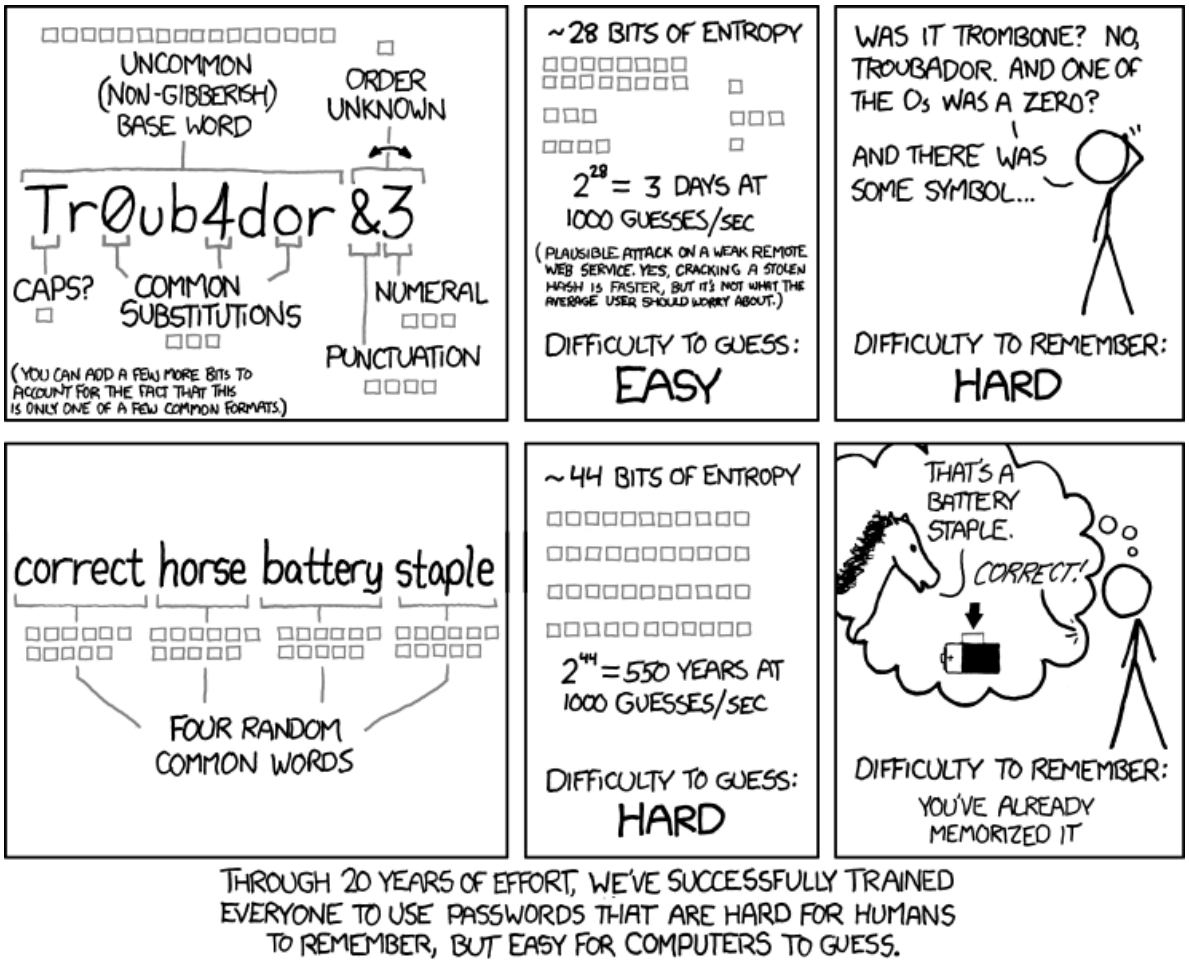
Exempel

- Vi behöver oberoende stokastiska variabler.
- Exempelvis:
  - längd,
  - antal och placering för varje teckenklass,
  - innehållet i varje tecken.
- Summan av entropierna för respektive del ger entropin för sannolikhetsfördelningen av lösenord.

Lösenord

- Om variablerna ej är oberoende får vi åtminstone en övre gräns – dock skulle vi vara mer intresserade av en nedre gräns.
- Denna metod används i [KSK+11], som ni ska läsa till seminariet.

Lösenord



Figur : xkcd:s serie om lösenordsstyrka. Bild: xkcd [xkc].

Lösenord

En förklaring av xkcd

- Vi har 1 miljon engelska ord: ger $\log 10^6 \approx 20$ bitar entropi.
- Vi kan ha inledande versal: ger 1 bit entropi.
- Vi har några vanliga substitutioner: uppskattningsvis 3 bitar entropi.
- Vi har specialtecken (ej substitution): uppskattningsvis 4 bitar entropi.
- Vi har siffror: $log 10 \approx 3$.
- Ordningen på specialtecknet och siffran: ger 1 bit entropi.
- Totalt 32 bitar entropi:
  - Tar minst 50 dagar med 1 000 gissningar per sekund.
  - Tar strax över en timme med 1 000 000 gissningar per sekund.

Lösenord

- Vi har 26 bokstäver, 10 siffror och (uppskattningsvis använder vi) 10 specialtecken.
- Detta ger totalt 72 möjliga tecken: ger $\log 72 \approx 6$ bitar entropi per tecken.
- Ett 10 tecken långt *slumpmässigt valt* lösenord ger således 60 bitar entropi.
- Tar 36 588 år mer 1 000 000 gissningar per sekund.

Lösenord

- Vi har 125 000 svenska ord: ger $\log 125\,000 \approx 17$ bitar entropi per ord.
- Ett lösenord med fyra slumpmässigt valda ord ger 68 bitar entropi.
- Tar 9 359 078 år att knäcka med 1 000 000 gissningar per sekund.

Lösenord

- Detta bygger på att vi väljer slumpmässigt.
- Vi är väldigt dåliga på slumpmässigt.
- Entropin kommer alltså att vara lägre då vi inte har en likformig sannolikhetsfördelning.

Lösenord

När de inte är slumpvisa

- [BS12] undersöker hur lingvistiken påverkar valet av lösenord bestående av flera ord.
- Finner att användarna inte väljer slumpmässiga ord utan föredrar att välja dem anpassade efter naturligt språk.
- Exempelvis ”correct horse battery staple” föredras framför ”horse correct battery staple” på grund av att det första alternativet är mer grammatiskt korrekt.

Lösenord

- [KRC06] gjorde en undersökning av hur användare skapar lösenord som är lätta att komma ihåg.
- Undersökte styrkan hos frasbaserade lösenord: skapas utifrån en mening och förkortas.
- Googles exempel ”To be or not to be, that is the question”[2] som ger lösenordet ”2bon2btitq”.
- I lösenordsdatabaser har just detta lösenord dykt upp . . .

[2] URL: `http://www.lightbluetouchpaper.org/2011/11/08/want-to-create-a-really-strong-password-dont-ask-google/`.

Lösenord

- Vi kan använda läckta lösenordsdatabaser för att uppskatta entropin för mänskligt valda lösenord.
- Vi borde också kunna uppskatta hur mycket information en given lösenordspolicy ger om lösenordet genom informationsvinst.

Översikt

Identifierande information

Exempel (Vad får vi ut mest information av?)

Får vi ut mest information av stjärntecken eller födelsedag?

$$-\sum_{\text{stjärntecken}} \frac{1}{12}\log\frac{1}{12} = \log 12 \approx 3.58$$

$$< -\sum_{\text{dagar på året}} \frac{1}{365}\log\frac{1}{365} = \log 365 \approx 8.51.$$

Notera

- Om jag ska gissa er födelsedag och får veta i vilket stjärntecken ni är födda gör jag en informationsvinst på ca 3.58 bitar.
- Då återstår ca 5 bitar att gissa, detta stämmer överens med $\log 30 \approx 4.91$.

Identifierande information

Hur mycket information behövs för att unikt identifiera en individ?

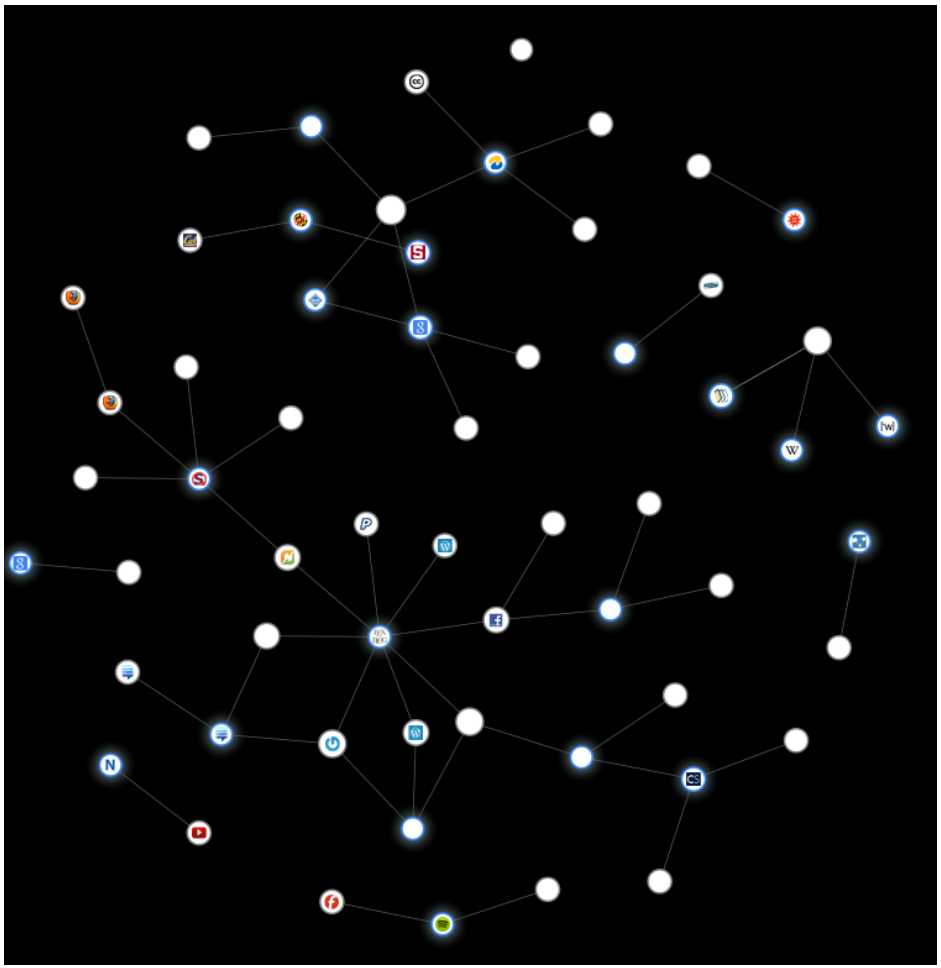
Identifierande information

- Det fanns $n = 6\,973\,738\,433$ människor på jorden, vid något tillfälle under 2011 enligt Världsbanken.
- Att identifiera en person kan ses som att välja med likformig sannolikhetsfördelning $1/n$.
- Följaktligen krävs $\log n = 32.7 \approx 33$ bitar information.

Identifierande information

- Electronic Frontier Foundation (EFF) genomförde en undersökning [Eck10a] hur mycket identifierande information en webbläsare delar med sig av.
- Går att testa sin webbläsare på `http://panopticlick.eff.org/`.
- Med min Firefox-läsare och alla tillägg uppskattar de att jag ger ifrån mig 21.45 bitar med identifierande information.
- De har genomfört 2 860 696 test totalt.

Identifierande information



Figur : Skärmdump från Collusion för Firefox över alla sidor som spårar mig med hjälp av denna information.

Översikt

Mittuniversitetet
MID SWEDEN UNIVERSITY

Kryptografi

- Används för att se hur mycket information vi får ut av en nyckel genom att se en kryptotext.
- Mer om detta under nästa föreläsning.

Referenser I

[Bos13] Daniel Bosk. Grundläggande lösenordsanalys. 2013.

[BS12] Joseph Bonneau och Ekaterina Shutova. Linguistic properties of multi-word passwords. I: *USEC*, 2012.

[Eck10a] Peter Eckersley. How unique is your browser? I: *Privacy Enhancing Technologies*, ss 1–18. Springer, 2010.

[Eck10b] Peter Eckersley. A primer on information theory and privacy, jan 2010.

[KRC06] Cynthia Kuo, Sasha Romanosky och Lorrie Faith Cranor. Human Selection of Mnemonic Phrase-based Passwords. Teknisk rapport 36, Institute of Software Research, 2006.

[KSK+11] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L. Mazurek, Lujo Bauer, Christin Nicolas, Lorrie Faith Cranor och Serge Egelman. Of passwords and people: Measuring the effect of password-composition policies. I: *CHI*, 2011.

Mittuniversitetet
MID SWEDEN UNIVERSITY

Referenser II

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, jul, oct 1948.

[Uel] Daniel Ueltschi. Chapter 6: Shannon entropy.

[xkc] xkcd. Password strength.