Indice

pag.

Prefazione IX

Introduzione 1

Capitolo 1

Big Data e protezione dei dati personali

1.1. Cosa è *Big* nei *Big Data*? 5
1.2. Internet e le *cose* 20
1.3. Perché la privacy è importante 23
1.4. I principi della protezione dei dati personali 29
1.5. Le nuove sfide per l'applicazione dei principi 31
1.6. Cosa è privacy by design 33
1.6.1. Il concetto di anonimizzazione e di dato anonimizzato 34
1.6.2. Il concetto di pseudonimizzazione e di dato pseudonimo 37

Capitolo 2

Anonimizzazione

2.1. Eventi aleatori 41
2.2. Il teorema di Bayes 48
2.3. Variabili aleatorie 51
2.4. Modelli per variabili aleatorie 60
2.4.1. Il modello di Bernoulli 60
2.4.2. Il modello binomiale 61
2.4.3. Il modello uniforme 62
2.4.4. Il modello gaussiano 63
2.4.5. Il modello di Laplace 64
2.5. Anonimizzazione per randomizzazione 67
2.5.1. Permutazione 67

pag.

2.5.2. Aggiunta di rumore 71
2.5.3. Differential privacy 73
2.5.4. I questionari polarizzati 75
2.6. Anonimizzazione per generalizzazione 81
2.6.1. Generalizzazione e classi di equivalenza 81
2.6.2. Il concetto di informazione ausiliaria 83
2.6.3. Diversi tipi di generalizzazione: k-anonymity, l-diversity, t-closeness 91
2.6.4. Una metodologia per la generalizzazione 101

Capitolo 3

Pseudonomizzazione

3.1. La crittografia 117
3.1.1. Crittografia a chiave simmetrica 120
3.1.2. Crittografia a chiave asimmetrica 125
3.2. L'architettura PKI 133
3.3. La previsione di comportamenti su dati pseudonimi 140
3.4. Una metodologia "*accountable*" per la previsione di comportamenti 141
3.5. Bigger is Better: un discorso sui volumi 148
3.6. Bigger is Better: un discorso sulla varietà 150
3.7. Bigger is Better: un discorso sulla velocità 154
3.8. La condivisione di dati 157
3.8.1. La condivisione bilaterale di dati privati: l'algoritmo di Diffie-Hellman 161
3.8.2. La condivisione multilaterale di dati privati 164

Capitolo 4

Pianificare la sicurezza

4.1. Sicurezza e privacy by design 171
4.2. Bigger is Better: un discorso sulla veracità (ossia sulla qualità) dei dati 174
4.3. Il teorema dell'incompletezza (o di Gödel): perché la sicurezza informatica è difficile 178
4.4. I programmi informatici visti come sistemi incompleti 181
4.5. Attacchi alle tecniche di anonimizzazione 187
4.6. Attacchi alla sicurezza 191
4.6.1. Attacchi alla confidenzialità 191
4.6.2. Attacchi all'integrità 196
4.6.3. Attacchi alla disponibilità 200

pag.

4.7. Investire in sicurezza 201
4.7.1. I costi della mancanza di sicurezza 201
4.7.2. Il modello di Gordon e Loeb 203
4.7.3. Investire nel cloud 207
4.8. Quali politiche per la sicurezza 210
4.8.1. Elementi di teoria dei giochi e il concetto di equilibrio di Nash 210
4.8.2. Free riders e condivisione dei danni 215

Bibliografia 219

Indice analitico 227

Box di approfondimento

Box 1.1. Il Machine Learning 15
Box 1.2. Limiti di un'informativa puramente testuale 32
Box 2.1. Il cigno nero 43
Box 2.2. La regola delle penalizzazioni appropriate 44
Box 2.3. Eventi composti e indipendenza 46
Box 2.4. Correlazione e probabilità condizionata 47
Box 2.5. La probabilità totale 50
Box 2.6. La distanza di Kullback-Leibler 56
Box 2.7. Valore atteso, varianza e deviazione standard di una variabile aleatoria 58
Box 2.8. La definizione di Differential privacy 74
Box 2.9. Il meccanismo del rumore laplaciano 75
Box 2.10. Come ottenere la stima corretta da questionari randomizzati 77
Box 2.11. La precisione della stima nei questionari con risposta randomizzata 78
Box 2.12. Come si costruisce un campione 86
Box 3.1. Un semplice esempio di crittografia a chiave simmetrica 122
Box 3.2. L'algoritmo RSA 128
Box 3.3. La generazione delle chiavi in RSA 129
Box 3.4. Un esempio di codifica e decodifica con RSA 130
Box 3.5. L'algoritmo di ElGamal 131
Box 3.6. La generazione delle chiavi con ElGamal 132
Box 3.7. Un esempio di codifica e decodifica con ElGamal 132
Box 3.8. La funzione di hash 136
Box 3.9. Due persone possono avere la stessa chiave crittografica? 137
Box 3.10. La collisione hash 139

pag.

Box 3.11. I classificatori "naive" di Bayes 147
Box 3.12. Il logaritmo discreto 162
Box 3.13. L'algoritmo di Diffie-Hellman 163
Box 3.14. Secure Multiparty Computation 167
Box 4.1. Gli standard sulla qualità dei dati ISO/IEC 25012 e ISO/IEC 25024 175
Box 4.2. Può un programma informatico essere etico? 186
Box 4.3. Investimenti e perdite nel modello di Gordon e Loeb 204