Geburtstags-Paradoxon

Dirk Fox

Gateway (Tor, Durchgang) ist ein Vermittlungscomputer, der zwei unterschiedliche, aber gleichartige Kommunikationssysteme verbindet. Dazu setzt er die Kommunikationsregeln (Protokolle) des einen in die des anderen um. So ermöglicht er den Teilnehmern beider Systeme, sich zu verständigen. In unserem „Gateway" werden Juristen technische und Technikern juristische Begriffe erläutert.

Unter dem sogenannten „Geburtstags-Paradoxon" wird eine auf den ersten Blick ausgesprochen verblüffende Erkenntnis verstanden, die in der Kryptologie eine wichtige Rolle spielt. Sie lässt sich in eine sehr anschauliche Frage kleiden, der das vermeintliche Paradoxon auch seinen Namen verdankt:

> *Wieviele Personen muss eine Gruppe umfassen, damit zwei von ihnen am selben Tag eines Jahres Geburtstag haben?*

Die Lösung ist ein verblüffend kleiner Wert: Es genügen 23 Personen.

Die mathematische Herleitung dieser Lösung ist vergleichsweise einfach: Nehmen wir dazu zunächst an, dass in einer Gruppe von n Personen (n<366) niemand am gleichen Tag Geburtstag haben soll wie eine andere Person dieser Gruppe. Besteht die Gruppe aus einer Person (n=1), so darf sie an einem beliebigen der 365 Tage des Jahres Geburtstag haben; die Wahrscheinlichkeit, dass dies zutrifft, ist 100% (also 1).

Kommt eine zweite Person hinzu (n=2), darf diese nur noch an einem der verbleibenden 364 Tage Geburtstag haben. Die Wahrscheinlichkeit, dass dies gilt, ist 364/365. Die n-te Person einer Gruppe darf nur noch an einem der noch nicht vergebenen 365-n+1 Tage Geburtstag haben; die Wahrscheinlichkeit dafür ist 366-n/365.

Will man nun die Wahrscheinlichkeit dafür bestimmen, dass alle n Personen einer Gruppe (n<366) an einem anderen Tag des Jahres Geburtstag haben (P_1), so muss man lediglich die Wahrscheinlichkeiten für die oben genannten Einzelereignisse miteinander multiplizieren:

$P_1(n)=1\cdot 364/365\cdot \ldots \cdot 366\text{-}n/365$.

Aus dieser Betrachtung lässt sich der gesuchte Wert, nämlich die Wahrscheinlichkeit, dass mindestens zwei Personen einer Gruppe aus n Personen am gleichen Tag Geburtstag haben (P_2), sehr leicht bestimmen, denn es gilt: $P_2=1\text{-}P_1$, d.h. die Wahrscheinlichkeit ist 100% minus der Wahrscheinlichkeit dafür, dass alle n Personen an einem anderen Tag Geburtstag haben.

Nun ist nur noch der kleinste Wert für n zu bestimmen, für den P_2>0,5 ist, d.h. die Wahrscheinlichkeit, dass zwei Personen der Gruppe am selben Tag Geburtstag haben, über 50% liegt. Dies gelingt sehr einfach durch Einsetzen verschiedener Werte für die Personenanzahl n:

$P_2(1) = 1\text{-}1 = 0$
$P_2(2) = 1\text{-}1\cdot 364/365 = 0{,}0027$ (≈ 0,3%)
...
$P_2(22) = 0{,}4757$ (≈ 48%)
$P_2(23) = 0{,}5073$ (≈ 51%)

Für 23 Personen liegt die Wahrscheinlichkeit, dass mindestens zwei am selben Tag Geburtstag haben, erstmals über 50%. Eine einfache Näherung für dieses Ergebnis ist, dass die gesuchte Zahl n etwas größer als die Quadratwurzel aus 365 (also ca. 20) ist. Dieses Resultat lässt sich wie folgt verallgemeinern:

> In einer Menge von n beliebigen Elementen, die je eine von insgesamt n^2 möglichen Eigenschaften besitzen, finden sich wahrscheinlich zwei, die dieselbe Eigenschaft besitzen.

Diese Erkenntnis hat weit reichende Konsequenzen für kryptografische Hashfunktionen, mit denen ein digital zu signierender „Fingerabdruck" zu einer Nachricht bestimmt wird, und *Message Authentication Codes* (MAC), die in vielen kryptographischen Protokollen die Integrität und Authentizität eines Datenpakets während der Übertragung sichern. Denn ähnlich wie das Geburtstags-Paradoxon lässt sich auch ein Fälschungsangriff auf eine digitale Signatur oder einen MAC formulieren:

> *Wieviele (beliebige) Nachrichten benötigt man, damit wahrscheinlich mindestens zwei von ihnen denselben Hashwert besitzten?*

Gelingt es, ein solches Nachrichtenpaar zu finden, so besitzen diese beiden Nachrichten mit dem selben Hashwert auch die selbe digitale Signatur. Wie beim Geburtstags-Paradoxon ist auch die Lösung für dieses Problem ein wesentlich kleinerer Wert als erwartet: Der Aufwand, eine Kollision für eine kryptografische Hashfunktion zu finden, liegt bei etwa $2^{k/2}$ Operationen; dabei gibt k die Länge des Hashwertes in bit an.

Um also sicher vor Angriffen mit dem Aufwand 2^{80} zu sein – das ist ein Aufwand der heute für symmetrische Kryptoverfahren als hinreichend sicher vor *brute force*-Angriffen gilt und einer Schlüssellänge von 80 bit entspricht – muss eine Hashfunktion daher eine 160 bit lange Ausgabe erzeugen. Dieser Anforderung genügen von den heute verbreiteten Hashfunktionen nur RIPEMD-160 [DoBP_96] und SHA-1 [NIST_95].

Auch für symmetrische Kryptoverfahren, die häufig zur Berechnung von MACs verwendet werden, hat diese Erkenntnis Bedeutung: Wird ein Kryptoverfahren mit einer Blocklänge von lediglich 64 bit (wie z.B. DES oder auch 3-DES) zur MAC-Berechnung eingesetzt, genügen 2^{32} Nachrichten, um eine Kollision zu finden.[1] MACs sollten daher ebenfalls mindestens eine Länge von 160 bit besitzen, um vor Fälschungsangriffen sicher zu sein.

Literatur

DaPr_89 Davies, Donald W.; Price, Wyn L.: *Security for Computer Networks*. 2. Auflage, John Wiley & Sons Ltd., Chichester 1989.

DoBP_96 Dobbertin, Hans, Bosselaers, Antoon; Preneel, Bart: *RIPEMD-160: A Stengthened Version of RIPEMD*. In: Gollmann, D. (Hrsg.): Proceedings of Fast Software Encryption 1996, LNCS 1039, Springer, Berlin 1996, S. 71-82.

NIST2_95 National Institute of Standards and Technology (NIST): *Secure Hash Standard (SHS-1)*. Federal Information Processing Standards Publication 180-1 (FIPS-PUB), 17.04.1995.

[1] In [DaPr_89] wird sehr anschaulich gezeigt, wie einfach es ist, aus einer Nachricht 2^m sinngleiche Varianten herzustellen: Es genügt, an m Stellen eines Textes jeweils für ein Wort ein sinngleiches anderes zu finden.