Diskret, leicht zu merken und massenhaft verfügbar

Die RWTH-Kundennummer ist eine einfache aber robuste Identitätsnummer

Wozu vergibt die RWTH eine Kundennummer? Wie bei allen Organisationen, die viele Personen als Kunden oder Mitglieder zu verwalten haben, reicht auch bei einer Hochschule der Name einer Person allein nicht aus, um ein Zeugnis auszustellen, ein Buch auszuleihen oder eine Klausuranmeldung vorzunehmen. Bei mehreren Zehntausend Studierenden, Alumni und Mitarbeitern ist die Gefahr zu groß, dem falschen "Thomas Lehmann" oder der falschen "Julia Schröder" eine Mahnung für ausgeliehene Bücher zu schicken oder sie mit einem vorzeitigen Diplomzeugnis für das falsche Fach zu beglücken. Eine Lösung wäre natürlich, den Namen immer nur in Kombination etwa mit der Anschrift, dem Geburtsort und dem Geburtsdatum zu verwenden. Allerdings sollen nicht bei jedem Verwaltungsvorgang all diese Informationen offengelegt werden müssen. Daher verwenden Organisationen in der Regel Kundennummern oder Mitgliedsnummern, um eine Person eindeutig zu identifizieren.

Leider nutzt dabei jede Organisation eine andere Methode, um die Nummern zuzuweisen. Das gilt nicht nur für die Vielzahl verschiedener Verwaltungen, wie etwa den Rentenversicherungen, Krankenversicherungen, Meldebehörden, Finanzämtern oder Automobilklubs, sondern bereits innerhalb einzelner Organisationen wie etwa der RWTH: Das Studierendensekretariat vergibt Matrikelnummern, die Personalstelle Personalnummern, die Bibliothek markiert Leseausweise mit einer Zahlenkombination und das Rechen- und Kommunikationszentrum vergibt Benutzerkennungen. Wie nicht anders zu erwarten, sind diese Nummern und Kennungen völlig unabhängig voneinander und die meisten betroffenen Personen benötigen mehrere solcher Zahlenkombinationen: Viele Studierende beispielsweise stehen als wissenschaftliche Hilfskräfte in einem Angestelltenverhältnis mit der RWTH und nutzen sowohl die Bibliothek als auch die Dienste des Rechen- und Kommunikationszentrums.

Im Rahmen der Einführung eines Identitätsmanagements an der RWTH wurde nun beschlossen, mit jeder Person, die irgendwelche Dienstleistungen der Hochschule in Anspruch nimmt, eine einheitliche Kundennummer zu assoziieren, mit Hilfe derer sich die notwendigen Informationen bei Bedarf aus den verschiedenen Datenbanken synthetisieren lassen. Die Kundennummer soll also als Datenbankschlüssel verwendet werden und deshalb die folgenden Eigenschaften haben:

Großer Nummernvorrat: Für die nächsten Jahrzehnte sollen ausreichend viele Kundennummern zur Verfügung stehen, ohne alte erneut verwenden zu müssen. Mit einem konservativen Ansatz ist von 10.000 neuen pro Jahr zu vergebenden Kundennummern auszugehen.
Anonymität: Aus der Kundennummer sollen keine personenbezogenen Informationen ablesbar sein. Zurzeit werden Matrikelnummern zum Beispiel fortlaufend vergeben, woraus sich das ungefähre Eintrittsdatum erkennen lässt.
Leichte Merkbarkeit: Die Kundennummer soll einprägsam sein und leicht als wiedererkannt werden.
Robustheit: Einfache Übermittlungs- oder Tippfehler sollen bemerkt werden, also keine gültige Kundennummer ergeben. Der Benutzer wird in diesem Fall einfach aufgefordert, die Zahlen neu einzugeben.

Zusammen mit Christian Bischof und Guido Bunsen vom Rechen- und Kommunikationszentrum wurde von uns folgender Vorschlag für eine einheitliche RWTH-Kundennummer, die sogenannte RWTH-ID, vorgelegt. Inzwischen hat sie Einzug in die Praxis gehalten: Bisher wurden fast 44.000 RWTH-IDs an Angehörige der Hochschule vergeben.

Die verwendeten Symbole

Zunächst ist festzulegen, aus welchen Symbolen die RWTH-ID gebildet werden soll. Beschränkt man sich auf die zehn Ziffern benötigt man 6 Ziffern für 10^6 = 1.000.000, also eine Million verschiedene Kundennummern und zusätzliche Stellen für Prüfziffern. Bildet man die Kundennummer jedoch aus den 26 Buchstaben des Alphabets unter Einbeziehung von Groß- und Kleinschreibung und den zehn Ziffern, so hat man 62 Symbole zur Verfügung. Damit können schon mit vier Symbolen 62^4 = 14.776.336, also fast 15 Millionen, Kundennummern gebildet werden. Für die RWTH-ID finden unter Einbeziehung von Ziffern und Großbuchstaben 32 verschiedene alphanumerische Symbole Verwendung, die in Bild 1 aufgeführt sind. Kleinbuchstaben und Symbole, die leicht verwechselt werden können, werden nicht benutzt: Die Symbole I und J unterscheiden sich nur wenig von 1, das Symbol O nur wenig von 0 und das Symbol V nur wenig von U. Deshalb kommen I, J, O und V nicht vor. Damit kann man mit vier Symbolen, 32^4 = 1.048.576, mehr als eine Million Kundennummern bilden und mit fünf Symbolen bereits 32^5 = 33.554.432. Das ist eine komfortable Anzahl.

00000	0		01000	8		10000	G		11000	R
00001	1		01001	9		10001	H		11001	S
00010	2		01010	A		10010	K		11010	T
00011	3		01011	B		10011	L		11011	U
00100	4		01100	C		10100	M		11100	W
00101	5		01101	D		10101	N		11101	X
00110	6		01110	E		10110	P		11110	Y
00111	7		01111	F		10111	Q		11111	Z

Bild 1: Übersetzung alphanumerischer Symbole in Binärfolgen.

Kommt noch ein Prüfsymbol dazu erhält man eine sechsstellige RWTH-ID. Um ihren Wiedererkennungswert zu steigern, wird eine einheitliche Schreibweise vereinbart, die immer verwendet wird, wenn eine RWTH-ID auf Bescheinigungen, Anträgen, Ausweiskarten oder Internet-Seiten ausgegeben wird: Es kommen immer genau sechs Symbole zum Einsatz, wobei zwei Gruppen von je drei Symbolen durch einen Bindestrich voneinander getrennt werden, etwa SL8-BRX. Man kennt ein ähnliches Verfahren auch von Kreditkarten, auf denen die 16-stellige Kreditkartennummer in Vierergruppen dargestellt wird.

Damit sind bereits die ersten drei der genannten Anforderungen an gute Kundennummern für die RWTH-ID erfüllt. Was aber ist mit der Robustheit? Wenn man das Prüfsymbol der RWTH-ID irgendwie zufällig wählt, kann es passieren, dass man einen Übermittlungs- oder Tippfehler womöglich nicht bemerkt und einen Kunden falsch identifiziert. Die bei Weitem häufigsten Tippfehler bei Tastatureingaben sind einzelne falsche Symbole, sogenannte “Einzelfehler”, und das

Vertauschen zweier benachbarter Symbole, sogenannte “Dreher”. Also kommt es darauf an, das Prüfsymbol so geschickt zu wählen, dass solche Fehler immer bemerkt werden. Hier kommt die Mathematik ins Spiel.

Die Kodierung

Zunächst werden die verwendeten Symbole in Folgen von fünf Binärziffern, also 0 oder 1, übersetzt, wie dies in Bild 1 angegeben ist. Es gibt gerade $2^5 = 32$ solcher Binärfolgen, was auch die Anzahl der ausgewählten Symbole erklärt. Eine RWTH-ID entspricht also einer Binärfolge der Länge 30; die alphanumerischen Symbole werden nur zur Ein- und Ausgabe benutzt. Für das Beispiel SL8-BRX erhält man etwa:

S	L	8	B	R	X
11001	10011	01000	01011	11000	11101

Polynome

Der Vorteil dieser Übersetzung ist, dass man mit diesen Binärfolgen sehr gut rechnen kann: Zunächst können die Binärziffern 0 und 1 addiert und multipliziert werden, so dass jeweils wieder 0 oder 1 herauskommt. Dies macht man einfach so, wie man es von den Zahlen 0 und 1 gewohnt ist, lediglich das Ergebnis der Addition von 1 und 1 ist nun eben 1 + 1 = 0. Fasst man Addition und Multiplikation als Bitoperationen auf, so sind dies übrigens gerade “exklusives Oder” und “Und”.

Eine Binärfolge $a_0, a_1, \ldots, a_d$ der Länge $d+1$ interpretiert man als formale Summe $f = a_0 x^0 + a_1 x^1 + a_2 x^2 + \ldots + a_d x^d$. Ausdrücke dieser Form werden als Polynome in der Variablen x bezeichnet. Nach dieser weiteren Übersetzung entspricht einer RWTH-ID also ein Polynom mit $d = 29$. Dabei gibt es zulässige und unzulässige Polynome, also solche, die zu einer gültigen RWTH-ID gehören, und solche, die dies nicht tun. Die Aufgabe besteht nun darin, festzulegen, welche Polynome zulässig sein sollen und welche nicht.

Die Summenschreibweise von Polynomen legt bereits nahe, wie man mit Polynomen sinnvoll rechnen, sie also addieren und multiplizieren kann. Für die a_i sind Addition und Multiplikation schon festgelegt und insgesamt rechnet man so, wie es die aus der Schule bekannten Rechengesetze befehlen: Für die Potenzen von x gilt das Potenzgesetz $x^{i+j} = x^i \cdot x^j$ und zwischen Addition und Multiplikation gilt das Distributivgesetz. Beim Rechnen mit Polynomen kommt es letztlich immer nur auf die zugrunde liegenden Binärfolgen an, deshalb können Addition und Multiplikation als Bitoperationen auf dem Rechner implementiert werden.

Die hier wichtigste Eigenschaft von Polynomen ist, dass sie analog zu den ganzen Zahlen eine Division mit Rest erlauben: “19 durch 7 ist 2 Rest 5”, das heißt es gilt die Gleichung $19 = 2 \cdot 7 + 5$. Die wesentliche Eigenschaft des eindeutig bestimmten Restes 5 ist, dass er kleiner als der Divisor 7 ist. Für Polynome kann man das ganz ähnlich so formulieren: Ist $f = a_0 x^0 + a_1 x^1 + a_2 x^2 + \ldots + a_d x^d$ mit $a_d = 1$, so heißt d der Grad von f, wobei sich der Grad eines Polynoms als Ersatz für das “kleiner” ganzer Zahlen erweist. Sind nämlich f und $g \neq 0$ Polynome, so gibt es eindeutig bestimmte Polynome q und r mit $f = q \cdot g + r$; wobei $r = 0$ oder der Grad von r kleiner als der Grad von g ist.

Mit einem der Division ganzer Zahlen völlig analogen Algorithmus kann man q und r berechnen,

wobei dieser Algorithmus leicht mit einem rückgekoppelten Schieberegister für die notwendigen Bitoperationen implementiert werden kann. Wie bei den ganzen Zahlen gibt Division mit Rest noch Anlass zu folgenden Sprechweisen: Hat f bei Division durch g den Rest $r = 0$, so sagt man g teile f analog zu "7 teilt 14"; und ist das Polynom 1 der einzige gemeinsame Teiler von f und g, so heißen f und g teilerfremd analog zu "11 und 7 sind teilerfremd".

Zurück zur RWTH-ID und zur Aufgabe festzulegen, welche Polynome zulässig sein sollen. Mit den eben eingeführten Sprechweisen geht das so: Man wählt ein geeignetes Polynom g, hält dieses fest und genau diejenigen Polynome sollen zulässig sein, die von g geteilt werden. Die möglichen RWTH-IDs entsprechen den Polynomen vom Grad kleiner 30 und wählt man g vom Grad k etwa, so kann man zeigen, dass es genau 2^{30-k} zulässige Polynome gibt. Um zu sehen, wie g genau gewählt werden muss, um eine robuste Kundennummer zu erhalten, werden Einzelfehler und Dreher in die neue Sprache übersetzt.

Einzelfehler und Dreher

Bei der Bestimmung eines geeigneten Polynoms g nutzt man aus, dass Tippfehler bei der Eingabe der RWTH-ID aufgrund der Kodierung der Symbole in Binärfolgen der Länge fünf nur in bestimmten Fenstern der Binärfolge auftauchen. Ein Einzelfehler verändert maximal fünf Binärziffern in den Positionen 1-5 oder 6-10 und so weiter aber nicht fünf Binärziffern etwa in den Positionen 3-7. Ein Dreher verändert analog maximal zehn Binärziffern in den Positionen 1-10 oder 6-15 und so weiter durch Vertauschen des jeweils ersten und zweiten Blocks der Länge fünf darin.

Ist f ein zulässiges Polynom, das heißt f wird von g geteilt, so liefert ein Tippfehler ein gestörtes Polynom f'. Der Tippfehler wird bemerkt, wenn f' nicht zulässig ist, also wenn f' nicht von g geteilt wird. Eine genaue Untersuchung der gestörten Polynome f', die von Einzelfehlern oder Drehern herkommen, zeigt, dass jeder Einzelfehler bemerkt wird, wenn g mindestens den Grad fünf hat und teilerfremd zum Polynom x ist und dass jeder Dreher bemerkt wird, wenn g teilerfremd zum Polynom $1+x^5$ ist.

Diese Analyse liefert also genaue Bedingungen an das zur Beschreibung gültiger RWTH-IDs benutzte Polynom g, so dass beim Eintippen einer RWTH-ID jeder Einzelfehler oder Dreher auch wirklich bemerkt wird. Da es möglichst viele zulässige Polynome geben soll, wählt man g von möglichst kleinem Grad mit den verlangten Eigenschaften, also vom Grad fünf . Man kann zeigen, dass es genau acht solcher Polynome vom Grad fünf gibt, darunter das sogenannte USB-5-Polynom $g=1+x^2+x^5$, das auch in anderen Bereichen der Informationstechnik eine Rolle spielt und für die RWTH-ID verwendet wird. Damit gibt es also $2^{30-5}=2^{25}=32^5$ zulässige Polynome, das heißt gültige RWTH-IDs.

Prüfung und Erzeugung

Die Prüfung einer RWTH-ID ist nun ganz einfach: Man übersetzt sie mittels Bild 1 in eine Binärfolge der Länge 30, diese in ein Polynom f vom Grad kleiner 30 und berechnet den Rest r der Division von f durch g. Nur wenn $r = 0$ ist, das heißt g teilt f, ist die untersuchte RWTH-ID gültig. Für das Beispiel SL8-BRX erhält man

$$f=\left(1+x+x^4\right)+\left(1+x^3+x^4\right)\cdot x^5+\left(x\right)\cdot x^{10}+\left(x+x^3+x^4\right)\cdot x^{15}+\left(1+x\right)\cdot x^{20}+\left(1+x+x^2+x^4\right)\cdot x^{25}$$

dabei gehören die eingeklammerten Ausdrücke zu den Blöcken der Länge fünf in der zugehörigen

Binärfolge und man kann nachrechnen, dass f von g geteilt wird.

Zur Erzeugung einer RWTH-ID geht man wie folgt vor: Man wählt eine beliebige Folge von fünf Symbolen aus Bild 1, die die letzten fünf Symbole einer gültigen RWTH-ID werden sollen. Diese übersetzt man in eine Binärfolge der Länge 25, dann in ein Polynom h vom Grad kleiner 25 und berechnet den Rest f der Division von $h \cdot x^5$ durch g. Dann ist $f = r + h \cdot x^5$ ein zulässiges Polynom vom Grad kleiner 30 und die daraus gebildete RWTH-ID besteht aus einem Prüfsymbol gefolgt von den gewählten fünf Symbolen. Zum Beispiel entspricht L8BRX dem Polynom

$$h = \left(1 + x^3 + x^4\right) + (x) \cdot x^5 + \left(x + x^3 + x^4\right) \cdot x^{10} + (1 + x) \cdot x^{15} + \left(1 + x + x^2 + x^4\right) \cdot x^{20} .$$

Division von $h \cdot x^5$ durch g ergibt den Rest $r = 1 + x + x^4$, der zur Binärfolge 11001, also zum Symbol S gehört.

Erweiterbarkeit

Sollte sich der Nummernvorrat doch irgendwann zur Neige gehen, so lässt sich die RWTH-ID übrigens auf beliebig lange Kundennummern erweitern, indem einfach die alten RWTH-IDs durch angehängte Nullen ergänzt werden. Damit geht keinerlei Verlust an Robustheit, sondern lediglich ein leichter Verlust von Anonymität einher, da es offensichtlich wird, wenn man zu den ersten paar Millionen Nutzern gehört!

Autor:

Privatdozent Dr. rer. nat Jürgen Müller ist Wissenschaftlicher Mitarbeiter am Lehrstuhl D für Mathematik