PGP word list

From Wikipedia, the free encyclopedia

The **PGP Word List** ('Pretty Good Privacy word list', also called a **biometric word list** for reasons explained below) is a list of words for conveying data bytes in a clear unambiguous way via a voice channel. They are analogous in purpose to the NATO phonetic alphabet used by pilots, except a longer list of words is used, each word corresponding to one of the 256 unique numeric byte values.

Contents

History and structure

The PGP Word List list was designed in 1995 by Patrick Juola, a computational linguist, and Philip Zimmermann, creator of PGP.[1][2] The words were carefully chosen for their phonetic distinctiveness, using genetic algorithms to select lists of words that had optimum separations in phoneme space. The candidate word lists were randomly drawn from Grady Ward's Moby Pronunciator list as raw material for the search, successively refined by the genetic algorithms. The automated search converged to an optimized solution in about 40 hours on a DEC Alpha, a particularly fast machine in that era.

The Zimmermann/Juola list was originally designed to be used in PGPfone, a secure VoIP application, to allow the two parties to verbally compare a short authentication string to detect a man-in-the-middle attack (MiTM). It was called a biometric word list because the authentication depended on the two human users recognizing each other's distinct voices as they read and compared the words over the voice channel, binding the identity of the speaker with the words, which helped protect against the MiTM attack. The list can be used in many other situations where a biometric binding of identity is not needed, so calling it a biometric word list may be imprecise. Later, it was used in PGP to compare and verify PGP public key fingerprints over a voice channel. This is known in PGP applications as the "biometric" representation. When it was applied to PGP, the list of words was further refined, with contributions by Jon Callas. More recently, it has been used in Zfone and the ZRTP protocol, the successor to PGPfone.

The list is actually composed of two lists, each containing 256 phonetically distinct words, in which each word represents a different byte value between 0 and 255. Two lists are used because reading aloud long random sequences of human words usually risks three kinds of errors: 1) transposition of two consecutive words, 2) duplicate words, or 3) omitted words. To detect all three kinds of errors, the two lists are used alternately for the even-offset bytes and the odd-offset bytes in the byte sequence. Each byte value is actually represented by two different words, depending on whether that byte appears at an even or an odd offset from the beginning of the byte sequence. The two lists are readily distinguished by the number of syllables; the even list has words of two syllables, the odd list has three. The two lists have a maximum word length of 9 and 11 letters, respectively. Using a two-list scheme was suggested by Zhahai Stewart.

Hex	Even Word	Odd Word
00	aardvark	adroitness
01	absurd	adviser
02	accrue	aftermath
03	acme	aggregate
04	adrift	alkali
05	adult	almighty
06	afflict	amulet
07	ahead	amusement
08	aimless	antenna
09	Algol	applicant
0A	allow	Apollo
0B	alone	armistice
0C	ammo	article
0D	ancient	asteroid
0E	apple	Atlantic
0F	artist	atmosphere
10	assume	autopsy
11	Athens	Babylon

Hex	Even Word	Odd Word
40	crackdown	Dakota
41	cranky	decadence
42	crowfoot	December
43	crucial	decimal
44	crumpled	designing
45	crusade	detector
46	cubic	detergent
47	dashboard	determine
48	deadbolt	dictator
49	deckhand	dinosaur
4A	dogsled	direction
4B	dragnet	disable
4C	drainage	disbelief
4D	dreadful	disruptive
4E	drifter	distortion
4F	dropper	document
50	drumbeat	embezzle
51	drunken	enchanting

Hex	Even Word	Odd Word
80	merit	intention
81	minnow	inventive
82	miser	Istanbul
83	Mohawk	Jamaica
84	mural	Jupiter
85	music	leprosy
86	necklace	letterhead
87	Neptune	liberty
88	newborn	maritime
89	nightbird	matchmaker
8A	Oakland	maverick
8B	obtuse	Medusa
8C	offload	megaton
8D	optic	microscope
8E	orca	microwave
8F	payday	midsummer
90	peachy	millionaire
91	pheasant	miracle

Hex	Even Word	Odd Word
C0	slowdown	recipe
C1	snapline	recover
C2	snapshot	repellent
C3	snowcap	replica
C4	snowslide	reproduce
C5	solo	resistor
C6	southward	responsive
C7	soybean	retraction
C8	spaniel	retrieval
C9	spearhead	retrospect
CA	spellbind	revenue
CB	spheroid	revival
CC	spigot	revolver
CD	spindle	sandalwood
CE	spyglass	sardonic
CF	stagehand	Saturday
D0	stagnate	savagery
D1	stairway	scavenger

12	atlas	backwater
13	Aztec	barbecue
14	baboon	belowground
15	backfield	bifocals
16	backward	bodyguard
17	banjo	bookseller
18	beaming	borderline
19	bedlamp	bottomless
1A	beehive	Bradbury
1B	beeswax	bravado
1C	befriend	Brazilian
1D	Belfast	breakaway
1E	berserk	Burlington
1F	billiard	businessman
20	bison	butterfat
21	blackjack	Camelot
22	blockade	candidate
23	blowtorch	cannonball
24	bluebird	Capricorn
25	bombast	caravan
26	bookshelf	caretaker
27	brackish	celebrate
28	breadline	cellulose
29	breakup	certify
2A	brickyard	chambermaid
2B	briefcase	Cherokee
2C	Burbank	Chicago
2D	button	clergyman
2E	buzzard	coherence
2F	cement	combustion
30	chairlift	commando
31	chatter	company
32	checkup	component
33	chisel	concurrent
34	choking	confidence
35	chopper	conformist
36	Christmas	congregate
37	clamshell	consensus
38	classic	consulting
39	classroom	corporate
3A	cleanup	corrosion
3B	clockwork	councilman
3C	cobra	crossover
3D	commence	crucifix
3E	concert	cumbersome
3F	cowbell	customer
52	Dupont	enrollment
53	dwelling	enterprise
54	eating	equation
55	edict	equipment
56	egghead	escapade
57	eightball	Eskimo
58	endorse	everyday
59	endow	examine
5A	enlist	existence
5B	erase	exodus
5C	escape	fascinate
5D	exceed	filament
5E	eyeglass	finicky
5F	eyetooth	forever
60	facial	fortitude
61	fallout	frequency
62	flagpole	gadgetry
63	flatfoot	Galveston
64	flytrap	getaway
65	fracture	glossary
66	framework	gossamer
67	freedom	graduate
68	frighten	gravity
69	gazelle	guitarist
6A	Geiger	hamburger
6B	glitter	Hamilton
6C	glucose	handiwork
6D	goggles	hazardous
6E	goldfish	headwaters
6F	gremlin	hemisphere
70	guidance	hesitate
71	hamlet	hideaway
72	highchair	holiness
73	hockey	hurricane
74	indoors	hydraulic
75	indulge	impartial
76	inverse	impetus
77	involve	inception
78	island	indigo
79	jawbone	inertia
7A	keyboard	infancy
7B	kickoff	inferno
7C	kiwi	informant
7D	klaxon	insincere
7E	locale	insurgent
7F	lockup	integrate
92	physique	misnomer
93	playhouse	molasses
94	Pluto	molecule
95	preclude	Montana
96	prefer	monument
97	preshrunk	mosquito
98	printer	narrative
99	prowler	nebula
9A	pupil	newsletter
9B	puppy	Norwegian
9C	python	October
9D	quadrant	Ohio
9E	quiver	onlooker
9F	quota	opulent
A0	ragtime	Orlando
A1	ratchet	outfielder
A2	rebirth	Pacific
A3	reform	pandemic
A4	regain	Pandora
A5	reindeer	paperweight
A6	rematch	paragon
A7	repay	paragraph
A8	retouch	paramount
A9	revenge	passenger
AA	reward	pedigree
AB	rhythm	Pegasus
AC	ribcage	penetrate
AD	ringbolt	perceptive
AE	robust	performance
AF	rocker	pharmacy
B0	ruffled	phonetic
B1	sailboat	photograph
B2	sawdust	pioneer
B3	scallion	pocketful
B4	scenic	politeness
B5	scorecard	positive
B6	Scotland	potato
B7	seabird	processor
B8	select	provincial
B9	sentence	proximate
BA	shadow	puberty
BB	shamrock	publisher
BC	showgirl	pyramid
BD	skullcap	quantity
BE	skydive	racketeer
BF	slingshot	rebellion
D2	standard	sensation
D3	stapler	sociable
D4	steamship	souvenir
D5	sterling	specialist
D6	stockman	speculate
D7	stopwatch	stethoscope
D8	stormy	stupendous
D9	sugar	supportive
DA	surmount	surrender
DB	suspense	suspicious
DC	sweatband	sympathy
DD	swelter	tambourine
DE	tactics	telephone
DF	talon	therapist
E0	tapeworm	tobacco
E1	tempest	tolerance
E2	tiger	tomorrow
E3	tissue	torpedo
E4	tonic	tradition
E5	topmost	travesty
E6	tracker	trombonist
E7	transit	truncated
E8	trauma	typewriter
E9	treadmill	ultimate
EA	Trojan	undaunted
EB	trouble	underfoot
EC	tumor	unicorn
ED	tunnel	unify
EE	tycoon	universe
EF	uncut	unravel
F0	unearth	upcoming
F1	unwind	vacancy
F2	uproot	vagabond
F3	upset	vertigo
F4	upshot	Virginia
F5	vapor	visitor
F6	village	vocalist
F7	virus	voyager
F8	Vulcan	warranty
F9	waffle	Waterloo
FA	wallet	whimsical
FB	watchword	Wichita
FC	wayside	Wilmington
FD	willow	Wyoming
FE	woodlark	yesteryear
FF	Zulu	Yucatan

Examples

Each byte in a bytestring is encoded as a single word. A sequence of bytes is rendered in network byte order, from left to right. For example, the leftmost (i.e. byte 0) is considered "even" and is encoded using the PGP Even Word table. The next byte to the right (i.e. byte 1) is considered "odd" and is encoded using the PGP Odd Word table. This process repeats until all bytes are encoded. Thus, "E582" produces "topmost Istanbul", whereas "82E5" produces "miser travesty".

A PGP public key fingerprint that displayed in hexadecimal as

```
E582 94F2 E9A2 2748 6E8B
061B 31CC 528F D7FA 8919
```

would display in PGP Words (the "biometric" fingerprint) as

```
topmost Istanbul Pluto vagabond
treadmill Pacific brackish dictator
goldfish Medusa afflict bravado
chatter revolver Dupont midsummer
stopwatch whimsical nightbird bottomless
```

The order of bytes in a bytestring is a topic discussed at length in computer science and engineering, and is beyond the scope of this article. This is often referred to as Endianness.

Other word lists for data

There are several other word lists for conveying data in a clear unambiguous way via a voice channel:

- the NATO phonetic alphabet maps individual letters and digits to individual words
- the S/KEY system maps 64 bit numbers to 6 short words of 1 to 4 characters each from a publicly accessible 2048-word dictionary. The same dictionary is used in RFC 2289.
- the Diceware system maps 5 base-6 random digits (almost 13 bits of entropy) to a word from a dictionary of 7,776 unique words.
- FIPS 181: Automated Password Generator converts random numbers into somewhat pronounceable "words".
- mnemonic encoding converts 32 bits of data into 3 words from a vocabulary of 1626 words.[3]

References

1. ^ Juola, Patrick; Zimmermann, Philip (1996). "Whole-Word Phonetic Distances and the PGPfone Alphabet" (http://www.mathcs.duq.edu/~juola/papers.d/icslp96.pdf) . *Proceedings of the International Conference of Spoken Language Processing (ICSLP-96)*. http://www.mathcs.duq.edu/~juola/papers.d/icslp96.pdf.
2. ^ Juola, Patrick (1996). "Isolated Word Confusion Metrics and the PGPfone Alphabet" (http://www.mathcs.duq.edu/~juola/papers.d/pgpfonenemlap.ps) . *Proceedings of New Methods in Language Processing 2* (Ankara, Turkey: Oxford University, Dept. of Experimental Psychology). http://www.mathcs.duq.edu/~juola/papers.d/pgpfonenemlap.ps.
3. ^ mnemonic encoding (http://www.tothink.com/mnemonic/) and updated code (http://github.com/singpolyma/mnemonicode)

Copyright

This material is copyrighted under a copyright owned by PGP Corporation. They have now granted a license under the GNU Free Documentation License. (per Jon Callas, CTO, CSO PGP Corporation, 4-Jan-2007)

Retrieved from "http://en.wikipedia.org/wiki/PGP_word_list"
Categories: Spelling alphabets | Military communications | Cryptography | PGP