# A Method for Approximating the Actual Key Size

E. P. Dawson [1] and H. M. Gustafson [2]

[1] Information Security Research Centre
Faculty of Information Technology
Telephone: 3864 1919
Fax: 3221 2384
Email: dawson@fit.qut.edu.au

[2] School of Mathematics
Faculty of Science
Telephone: 3864 2313
Fax: 3864 2310
Email: h.gustafson@qut.edu.au

Queensland University of Technology
Gardens Point
GPO Box 2434 Brisbane 4001
Queensland Australia

## Summary

The actual level of security of a particular encryption algorithm may be greatly reduced if there is redundancy in the method used to generate the key. It is shown how to use the number of repetitions in a sample of keys to approximate the actual key size.

## Extended Abstract

Cryptographic keys may be generated using hardware devices, or by using software that includes pseudo-random algorithms ([1], [3]). It is important that the hardware or software devices that are employed to generate these keys, be tested periodically to ensure the randomness of the output blocks. An attacker may find it easier to determine the key used in generation, by searching a small set of possible keys, than to investigate any possible weaknesses of the cipher used in encryption. For example, in late 1995 a potentially serious flaw was discovered in the Netscape Navigator [4] that reduced the size of the key space to far below the 40 or 128 bits allowed. Hence it is of fundamental importance that the method of generating the initialization keys, for the cryptographic algorithm, is itself secure.

In a sample of $K$ randomly generated $k$-bit keys, then each of the $N = 2^k$ possible binary $k$-bit patterns is equally likely to result for any one key. The distribution of the number of patterns that appear an exact number of times, has been extensively studied as $K, N \to \infty$ [2]. As the key length increases and $N \to \infty$, then for a sample

| $1-\alpha$ | $\lambda_L$ | $\lambda_U$ |
|---|---|---|
| 0.95 | $r - 1.96\sqrt{r}$ | $r + 1.96\sqrt{r}$ |
| 0.99 | $r - 2.5758\sqrt{r}$ | $r + 2.5758\sqrt{r}$ |

Table 1: Probability interval for the actual value of $\lambda$.

of size $K \to \infty$, where $\frac{K}{N} \to 0$, the distribution of the number of key repetitions in a the sample approaches a Poisson distribution with a mean of

$$\lambda = K - N(1 - e^{-\frac{K}{N}}). \tag{1}$$

In the event of the number of key repetitions, $r$, being much larger than expected, in a sample of $K$ keys, then it is likely that the *actual key space* (actual number of possible keys that may be obtained) has been reduced. A method for obtaining a probability interval for this *actual key space* will now be given.

For large values of $r$, (say $r > 20$), the probability distribution $\lambda$ may be approximated by the normal distribution with both mean and variance equal to $r$. Hence a $(1-\alpha)$ probability interval for the true value of $\lambda$ may be approximated using

$$r \ \pm \ z_{\frac{\alpha}{2}} \times \sqrt{r}, \tag{2}$$

where $z_{\frac{\alpha}{2}}$ represents a percentage point of the distribution of the standard normal variable, $z$, such that $Pr(z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$. Table 1 gives the lower, $\lambda_L$, and upper, $\lambda_U$, endpoints for the most commonly used probability intervals. Given values of $K$ and $\lambda$, then $N$ may be solved from (1) using

$$N(1 - e^{-\frac{K}{N}}) = K - \lambda. \tag{3}$$

The endpoints of the probability interval for $\lambda$ obtained from (2) are substituted into (3) to obtain a probability interval for the *actual key space*, $N$,based on the number of repetitions, $r$, in the sample. The endpoints of the probability interval may be determined using mathematical software packages or iterative methods. A simple approximation for $N$ in (3) is $N \approx \frac{K^2}{2\lambda}$. If the number of repetitions is more than expected for the sample of keys tested, then this calculation will result in a probability interval for the size of the *actual key space* that is available for the generator.

Table 2 contains examples of the endpoints for both 0.95 and 0.99 probability intervals, in samples of one million keys, when the number of repetitions is 30, 50 or 100. The sample size chosen, of $K = 10^6 \approx 2^{20}$ key blocks, would satisfy the requirements of the Poisson approximation in (1) for $k \geq 40$. Since $2^{32} \approx 4.29 \times 10^9$ and $2^{35} \approx 3.44 \times 10^{10}$, then it can be seen from Table 2 that 100 repetitions in samples of $K = 10^6$ key blocks would indicate that the key space is reduced to approximately $2^{33}$ possibilities, equivalent to a block length of at most 33 independent bits, with a probability exceeding 0.995.

| Repetitions r | Probability Interval | | | |
|---|---|---|---|---|
| | 0.95 | | 0.99 | |
| | Lower | Upper | Lower | Upper |
| 30 | $1.2274 \times 10^{10}$ | $2.5954 \times 10^{10}$ | $1.1335 \times 10^{10}$ | $3.1463 \times 10^{10}$ |
| 50 | $7.8294 \times 10^{9}$ | $1.3834 \times 10^{10}$ | $7.3296 \times 10^{9}$ | $1.5730 \times 10^{10}$ |
| 100 | $4.1803 \times 10^{9}$ | $6.2186 \times 10^{9}$ | $3.9756 \times 10^{9}$ | $6.7344 \times 10^{9}$ |

Table 2: Probability intervals for *actual key space* from samples of $10^6$ keys.

# References


[1] D. Eastlake, S. Crocker, and J. Schiller, *RFC1750: Randomness Recommendations for Security*, Internet Activities Board, December 1994.

[2] V. F. Kolchin, The Speed of Convergence to Limit Distributions in the Classical Ball Problem, *Theory of Probability and its Applications* **11**, 1966, 128-140.

[3] A. J. Menzies, P. C. van Oorschot and S. A. Vanstone, *CRC Handbook of Applied Cryptography*, CRC Press, 1996.

[4] M. E. Zurko, LISTWATCH: WWW-Security, *Electronic CIPHER*, Newsletter of the IEEE Computer Society's TC on Security and Privacy, Carl Landwehr ed., Electronic Issue 10, November 1, 1995.