Performance and Statistical Analysis of Stream Ciphers in GSM Communications

Nagendar Yerukala, V Kamakshi Prasad, and Allam Apparao

Original scientific paper

Abstract—Analysis of stream cipher can be done in two ways (1). Implementation aspects of the design (2). Statistical weakness of the keystream generated by stream cipher. Our research work mainly focused on statistical analysis of stream ciphers used in GSM communications. For a stream cipher to be secure, the keystream generated by it should be uniformly random (i.e., the keystream should follow a Bernoulli distribution with parameter 1/2). Statistical tests check whether the given sequence follows a certain probability distribution. In this paper, statistical analysis and correlation tests have been applied to various stream ciphers used in GSM 2G, 3G, 4G, and 5G communications to check for any weaknesses, which have not been done previously. The sequences output by these ciphers are checked for randomness using the statistical tests defined by the NIST Test Suite [6]. Furthermore, it should also be not possible to derive any information about the secret key and the initial state of the cipher from the keystream. Therefore, additional statistical tests based on properties like correlation between keystream and key, and correlation between keystream and Initial Vector (IV) described in [2] are also performed. Performance analysis of the ciphers has also been done and the results have been tabulated. Almost all the ciphers pass the tests in the NIST test suite with 99% confidence level. For A5/3 stream cipher, the correlation between the keystream and key is high and correlation between the keystream and IV is low when compared to other ciphers in the A5 family.

Index Terms—Stream Ciphers, NIST, Statistical Randomness Testing, GSM, Correlation, Encryption, Keystream.

I. INTRODUCTION

STREAM CIPHER is an encryption algorithm which takes a short secret key and produces a keystream to be xored with the clear text message. This is a key dependent algorithm. Stream ciphers are widely used in telecommunication applications, to provide privacy for communication. Stream ciphers are efficient and easy to implement in both hardware and software. They can be used to encrypt / decrypt data in real time such as encrypting the DVD content, playing back the content in real time after decryption [15], [16]. However, there are no sufficient details about the security of stream ciphers in the literature [14], [25]-[27].

A stream cipher can be viewed as a finite state machine which takes two inputs: secret key K and a publicly known Initial Vector (IV) (optional), and generates a pseudo-random keystream sequence.

A stream cipher can be attacked either at: Key / Initial Vector (IV) initialization procedure or at Keystream generation procedure. (1). Key / Initial Vector (IV) initialization procedure: To check whether the key / IV initialization procedure of GSM ciphers is secure or not, we performed the tests based on the correlation between key and keystream and the correlation between IV and keystream. From our experiments, it has been observed that there is no flaw in the key and IV initialization procedure of GSM ciphers (A5 family). All the cryptanalytic attacks (i.e., Rainbow Table attack, Brute Force, Time-Memory Trade-Off (TMTO) attacks, Guess and Determine attack, etc.) have focused on recovering the internal / initial state of a stream cipher given the keystream. Not many of them have focused on targeting the key / IV initialization procedure. (2). Keystream generation procedure: Statistical tests performed on the keystream usually can not detect any serious weaknesses of a stream cipher. However, the keystream should pass all the statistical tests. Otherwise, we can build a distinguisher which can distinguish the keystream from a random-looking sequence, which can then be used in a cryptanalytic attack.

The security requirements of a stream cipher are: 1. The keystream should be indistinguishable from a uniformly random sequence. 2. It should not be possible to derive any information about the secret key and the initial state of the stream cipher from the keystream.

Statistical tests are performed to know whether the keystream of a cipher can be distinguished from a truly random sequence. These tests do not take the inner workings of the cipher into account. They define test-statistics (cipher-independent) that can be computed from the keystream. The probability distribution of the test-statistic under the assumption of randomness is already known. If the probability of test-statistic value computed from the keystream is below a certain threshold, then the keystream is distinguishable from a truly random sequence. Otherwise, it is not.

If correlation between keystream and key or correlation between keystream and IV is away from 1/2 then it is possible to guess the key and IV bits, given the keystream, with probability greater than 1/2.

A. Contribution and Related Work

In this paper, we perform a detailed statistical analysis of various stream ciphers used in GSM 2G, 3G, 4G and 5G communications to check for any weaknesses, which have...
not been done previously. In [2], similar statistical analysis has been applied to synchronous stream ciphers submitted to ECRYPT [3].

This paper is organized as follows. In section II, various stream ciphers used in GSM communication are described. In section III, the statistical test based on correlation between the Keystream and Key & the statistical test based on correlation between the Keystream and IV [2] are described and also the results of applying these tests on the GSM ciphers are provided. The results of applying NIST test suite on all the ciphers are presented in section VII. Finally, the conclusion is given in section V.

Abbreviations:
- GSM: Global System for Mobile Communications
- GPRS: General Packet Radio Services (2G & 3G)
- UMTS: Universal Mobile Telecommunications Service (3G)
- LTE: Long-Term Evolution (4G)
- EC-GSM-IoT: Extended Coverage GSM for IoT
- 5G-NR: 5G New Radio is the Fifth-Generation wireless air interface

II. STREAM CIPHERS IN GSM COMMUNICATIONS

Definition 1 (Synchronous Stream Cipher [1]): A synchronous stream cipher with key-space \( \{0,1\}^n \) and IV-space \( \{0,1\}^m \) consists of:
- An internal state of \( s \) bits,
- A state initialization function \( \text{Int}: \{0,1\}^s \times \{0,1\}^n \rightarrow \{0,1\}^s \),
- A state update function \( \text{Udt}: \{0,1\}^s \rightarrow \{0,1\}^s \), and
- An output function \( \text{Opt}: \{0,1\}^s \rightarrow \{0,1\}^m \).

It takes a key-IV pair \( (K, IV) \in \{0,1\}^s \times \{0,1\}^n \) as input and outputs a key stream \( z = (z_1, z_2, \ldots) \), where \( z_i \in \{0,1\}^m \), as follows:

1. compute initial state \( S_0 = \text{Int}(K, IV) \in \{0,1\}^s \),
2. for \( i = 1, 2, \ldots \)
   2a) output \( z_i = \text{Opt}(S_{i-1}) \),
   2b) update state \( S_i = \text{Udt}(S_{i-1}) \).

A stream cipher is called word-based if \( m > 1 \). For \( m = 1 \), it is usually referred to as bit-oriented stream cipher. Both State

\[
(K, IV) \xrightarrow{\text{Int}} S_1 \quad \xrightarrow{\text{Udt}} S_2 \quad \xrightarrow{\text{Udt}} S_3 \quad \xrightarrow{\text{Udt}} \ldots \\
\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \\
(z_1, z_2, z_3, \ldots)
\]

Fig. 1. Stream Cipher

update function \( \text{Udt} \) and output function \( \text{Opt} \) must satisfy the following criteria for the cipher to be considered secure:
- It should not be possible to guess the internal state given the output of the \( \text{Opt} \) function.
- The \( \text{Udt} \) function must ensure that the stream cipher has high period, no matter whatever the internal state the stream cipher is initially in.

A. Stream Ciphers in Communication

\( A5/1 \) [4], [9]-[14] is an LFSR based stream cipher used in GSM communications and has been developed primarily for European markets. It takes as input 22 bit IV(frame number) and a 64 bit key. It has three LFSRs which are irregularly clocked.

\( A5/2 \) [4], [14], [17]-[19] is an LFSR based stream cipher used in GSM communications and has been developed primarily for European markets. It takes as input 22 bit IV(frame number) and a 64 bit key. It has four LFSRs which are irregularly clocked.

\( A5/3 \) [4], [14], [20]-[24] uses KASUMI block cipher in Output Feedback (OFB) mode. GSM-3G technology uses A5/3 with 64 bit key and GPRS(GEA3 algorithm) and UMTS(UEA1 algorithm) use A5/3 with 128 bit key for encryption.

\( SNOW3G \) [4], [25]-[28] is an LFSR based and has a finite state machine. It has been used by the various standards both for encryption and integrity checking as shown in the Table I. ZUC [4], [25]-[27], [29] is word-based LFSR and uses a non-linear function for producing output. 4G-LTE uses ZUC for encryption(128-EEA1 algorithm) and also for integrity(128-EIA3 algorithm:32 bit MAC based ZUC). 5G New Radio (5G-NR) uses ZUC for encryption(NIA3 algorithm) and also for integrity (NIA3 algorithm:32 bit MAC based ZUC).

The basic parameters and components of these ciphers are given in the Table II. The system configuration on which these algorithms have been run to calculate the throughput is: Ubuntu 14.04 LTS with intel core i7CPU 860@2.80GHZ x8 and OS type: 64-bit. Our study can be useful to the researchers to check the throughput of their design before implementing it on Hardware.

III. STATISTICAL TESTS BASED ON CORRELATION

To check whether the key / IV initialization procedure of stream ciphers used in GSM is secure or not, the correlation between key and keystream and the correlation between IV and keystream [2] have been performed on stream ciphers used in GSM.

A. Correlation Test between Keystream and Key

The aim is to compute the correlation between the keystream and key. If there is any correlation between the keystream and key then either the attacker can recover the secret key using the keystream or it may reduce the search space of brute force attack for finding the secret key. If a stream cipher fails this test, then it is necessary to revise the key initialization process.
Let the key size be \( l \) and the IV size be \( m \). The test procedure is as follows [2]:

- Fix IV as a zero vector
- For \( i = 1 \) to \( 1048576 (= 2^{20}) \) do
  - Generate a random key \( K_i \)
  - Run the cipher \( SC(K_i, IV) \) and generate the keystream \( Z_i \) of length \( l \) for each random key \( K_i \)
  - Find Hamming Weight \( W_i = HW(Z_i \oplus K_i) \), where \( 0 \leq W_i \leq l \). \( W_i \)'s are called observed frequencies
- If the probability distribution of these hamming weights is Binomial\((n, p)\) where \( n = l \) and \( p = 1/2 \), then the stream cipher is considered secure
- Compute probabilities of these weights using the binomial distribution (expected frequencies)
- Group these weights into classes so that each class has approximately the same probability
- Apply \( \chi^2 \)-Goodness of Fit Test
- Return \( p \)-value

If the hamming weight of a sequence obtained by xoring the key and keystream is too low/high then it indicates that both key and keystream are correlated.

This test is applied on the GSM stream ciphers and results are shown in Table III and Table IV. It is observed that all \( p \)-values are greater than or equal to 0.01. Hence these stream ciphers are secure according to this test.

### B. Correlation Test between Keystream and IV

The aim is to compute the correlation between the keystream and Initial Vector (IV). If there is any correlation between the keystream and IV then the attacker can generate (part of) the keystream without having the knowledge of key.

### IV. Randomness Tests

Randomness tests check the degree of approximation of the given binary sequence to a truly random sequence. This
TABLE V
CORRELATION BETWEEN KEYSRTEAM AND IV OF A5 FAMILY (64-BIT KEY CIPHERS).

| Class (Group) | Observed | Expected | p-value |
|---------------|----------|----------|---------|
| A5/1          | 260736   | 261289   | 261832  |
| A5/2          | 154162   | 154595   | 154155  |
| A5/3          | 168834   | 168477   | 167981  |
| A5/1          | 261733.532 | 0.149    | 0.141   |
| A5/2          | 154172.421 | 0.892    | 0.4239  |
| A5/3          | 168188.095 | 0.091    | 0.031   |

TABLE VI
CORRELATION BETWEEN KEYSRTEAM AND IV OF ZUC AND SNOW3G (128-BIT KEY CIPHERS).

| Class (Group) | ZUC Observed | ZUC Expected | SNOW3G Observed | SNOW3G Expected | p-value |
|---------------|--------------|--------------|-----------------|-----------------|---------|
| A5/1          | 165433       | 165074       | 165467.9145     | 0.2115          |
| A5/2          | 229504       | 229573       | 230035.8102     | 0.4239          |
| A5/3          | 209526       | 208993       | 208992.5506     | 0.149           |
| A5/1          | 166040       | 165945       | 165467.9144     | 0.892           |

is usually done through the use of a test-statistic. This test-statistic follows a certain probability distribution $P_0$ for a truly random sequence. The test-statistic value $v_0$ is computed for the given binary sequence to be tested and the probability that the test-statistic assumes the value $v_0$ under the probability distribution $P_0$ is computed (This probability is known as $p$-value). If the $p$-value is greater than certain threshold, then the binary sequence is considered to be truly random. Otherwise, it is not.

The test suites which are available in the literature for performing randomness tests are: NIST [6], DIEHARD [7], and Federal Information Processing Standard tests (FIPS) [8].

National Institute of Standards and Technology (NIST) has developed 15 randomness tests to test the randomness of the sequences. Every statistical test defined in the NIST test suite takes the sequence output by a Pseudo Random Number Generator (PRNG) as input and outputs whether sequence is random (according to property specified by the test) or not. If any of the sequence output by a PRNG is not random (based on some statistical test), we say that it is not a PRNG. Since we want to check whether the stream ciphers behave as pseudo-random bit generators (PRBG), we apply statistical tests on the sequence output by the stream cipher. Key stream of size 1048576 has been generated from each stream cipher mentioned in section II. The following inputs are given to stream ciphers for generating the keystream.

- A5/1
  - Key: 0x1223456789ABCDEF
  - IV: 0x134
- A5/2
  - Key: 0x1223456789ABCDEF
  - IV: 0x134
- A5/3
  - Key: 0x704FBD4EFA0BC0F3
  - IV: 0x1D8AF5
- SNOW3G
  - Key: 0xaaaaaaa 0x1234567890abcdef00b
  - IV: 0x01001000100010001000100010001000
- ZUC
  - Key: 0x41 0x42 0x43 0x44 0x45 0x46 0x47 0x48 0x49 0x4a 0x4b 0x4c 0x4d 0x4e 0x4f 0x50
  - IV: 0x01 0x02 0x03 0x04 0x05 0x06 0x07 0x08 0x09 0x0a 0x0b 0x0c 0x0d 0x0e 0x0f 0x10

A p-value is determined for each test in the NIST test suite. If $p = 1$ then the given keystream is said to be perfectly random sequence and if $p = 0$ then the given keystream is said to be non-random sequence. The significance level $\alpha$ for the tests is selected to be 0.01. If $p \geq \alpha = 0.01$ then with
confidence level 99%, we can say that the generated keystream is random and if \( p < \alpha = 0.01 \) then the keystream is said to be non-random with confidence level of 99%.

The NIST test suite [2], [5], [6] is applied on each generated keystream and results are tabulated in the Table VII. The \( p \)-values of all the ciphers for each test have been depicted graphically from Fig. 2 to Fig. 15. Almost all the ciphers pass the tests in NIST test suite with 99% confidence level. From the figures Fig. 2 to Fig. 15, one can observe that \( p \)-values are higher for some cipher which means that it possesses statistical properties similar to that of a random sequence when compared to other ciphers. For example, the frequency test checks the proportion of number of 1s and 0s in the given sequence and checks the difference falls within the limit of randomness. According to this test, from the Fig. 2, the keystream generated by A5/2 has highest \( p \)-value 0.811663 and ZUC has lowest \( p \)-value 0.096491. From the Table VII, keystream generated by ZUC has \( p \)-value 1.000000 in 6th of 18 sub-tests in Random Excursions Variant Test. Non-overlapping template matching test has 148 sub-tests. It searches for the pre-specified m-bit string in a given sequence of keystream bits and examine whether the number of such occurrences are within the statistical limit of a sequence. None of the ciphers passed all the 148 sub-tests. Keystreams produced by A5/1, A5/2, A5/3, ZUC, and SNOW3G have passed 129, 132, 136, 136, and 134 sub-tests respectively out of 148 tests.
TABLE VII
NIST TEST SUITE RESULTS.

| Test                                      | p-values          |
|-------------------------------------------|-------------------|
| A5/1 | A5/2 | A5/3 | ZUC | SNOW3G |
|-------------------------------------------|-------------------|
| 1. Frequency (Monobit) Test               | 0.778519 | 0.811663 | 0.688869 | 0.096491 | 0.781516 |
| 2. Frequency Test within a Block          | 0.778519 | 0.774172 | 0.615560 | 0.835373 | 0.780045 |
| 3. Runs Test                              | 0.796492 | 0.393404 | 0.684444 | 0.966320 | 0.480715 |
| 4. Test for the Longest Run of Ones in a Block | 0.667237 | 0.399498 | 0.060836 | 0.270085 | 0.574408 |
| 5. Binary Matrix Rank Test                | 0.534932 | 0.385795 | 0.213581 | 0.423972 | 0.295905 |
| 6. Discrete Fourier Transform (Spectral) Test | 0.538712 | 0.358795 | 0.611998 | 0.847917 | 0.594507 |
| 7. Non-overlapping Template Matching Test (No. of SUCCESSES out of 148 Sub-Tests) | 129 | 132 | 136 | 136 | 134 |
| 8. Overlapping Template Matching Test     | 0.257392 | 0.444027 | 0.674437 | 0.490965 | 0.223020 |
| 9. Maurer’s “Universal Statistical” Test  | 0.150609 | 0.761597 | 0.009726 | 0.915651 | 0.103432 |
| 10. Linear Complexity Test                | 0.379353 | 0.276610 | 0.100078 | 0.766208 | 0.328045 |
| 11. Serial Test                           | 0.666495 | 0.048873 | 0.790829 | 0.825544 | 0.540077 |
| 12. Approximate Entropy Test              | 0.665547 | 0.343203 | 0.655581 | 0.529153 | 0.136982 |
| 13. Cumulative Sums (Cusum) Test          | 0.919818 | 0.662897 | 0.290191 | 0.137821 | 0.899221 |
| 14. Random Excursions Test                | 0.656820 | 0.814566 | 0.569050 | 0.010799 | 0.850922 |
| 15. Random Excursions Variant Test       | 0.712781 | 0.051011 | 0.547507 | 0.096591 | 0.823987 |
| 16. Discrete Fourier Transform (Spectral) Test | 0.947543 | 0.947543 | 0.947543 | 0.947543 | 0.947543 |

Fig. 12. Serial Test-1

Fig. 13. Approximate Entropy Test

Fig. 14. Cumulative Sums Test (Forward)

Fig. 15. Cumulative Sums Test (Reverse)
V. CONCLUSION

In this paper, performance analysis, randomness tests, correlation between keystream and key, and correlation between keystream and IV are applied on GSM Stream Ciphers and results are tabulated and compared.

From the comparison of tests, one can observe that it is very difficult to find the relationship between the design structure and randomness p-values.

If a stream cipher fails a particular statistical test then with 100% probability, it can be distinguished from a truly random sequence. However, if a stream cipher passes a particular statistical test then with 100% probability it can not be proved that it is indistinguishable from a truly random sequence. This is because there are infinitely many statistical tests that can be applied to a keystream. However, passing all of them does not ensure that stream cipher is secure against all cryptanalytic attacks. Therefore, the statistical analysis that is performed in this paper is a necessary but not sufficient condition to prove that a stream cipher is secure. However, if it fails any of them, a cryptanalytic attack can be launched against the stream cipher.

Our work focused on statistical analysis of pseudo random keystream sequence generated by the stream ciphers used in GSM communications. Based on our approach one can understand that: (1). If there is a correlation between key and keystream then a key initialization process should be revised. (2). If there is a correlation between Initial vector and key stream then IV initialization process should be revised. There are other correlation tests in the literature that have to be applied on GSM ciphers as a future work.

REFERENCES

[1] Schneier, Bruce. (1995). Applied Cryptography: Protocols, Algorithms, and Source Code in C, 2nd Edition. John Wiley & Sons, Inc., New York, USA.

[2] Sönmez Turan, M., DoGanaksoy, A. and Çalık, Ç. (2006). Statistical analysis of synchronous stream ciphers, In: SASC 2006: Stream Ciphers Revisited, Leuven, Belgium.

[3] https://www.ecrypt.eu.org/stream/index.html

[4] https://gdelugre.github.io/2018/05/10/3gpp-ota-security-evolution/# encryption

[5] Eljadi. Fardous, and Fakhin Al-Shaikhli. Imad. (2015), Statistical Analysis of the eSTREAM Competition Winners, pp.80-85, https://doi.org/10.1109/ACSAT.2015.43

[6] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray, and S. Vo. (2008), A statistical test suite for random and pseudorandom number generators for cryptographic applications, National Institute of Standards and Technology (NIST), special publication 800-22.

[7] G. Marsaglia. (1995), DIEHARD Statistical Tests. http://stat.fsu.edu/~geo/diehard.html

[8] NIST. (2001), Federal Information Processing Standards (FIPS) Publications: FIPS 140-2, Security Requirements for Cryptographic Modules, http://csrc.nist.gov/publications/PubsFIPS.html#140-2.

[9] J. Golic. (1997), Cryptanalysis of Alleged A5 Stream Cipher, proceedings of EUROCRYPT’97, LNCS 1233,pp.239-255, Springer-Verlag.

[10] Maximov Alexander, Thomas Johansson, and Steve Babbage. (2004), An Improved Correlation Attack on A5/1, Selected Areas in Cryptography, pp.1-18.

[11] E. Barkan, E. Biham, and N. Keller. (2003), Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication, Advances in Cryptology – CRYPTO 2003, Lecture Notes in Computer Science 2729, pp. 600-616, Springer-Verlag.

[12] Biryukov, Alex, Shamir, Adi, and Wagner, David. (2001), Real Time Cryptanalysis of A5/1 on a PC, Fast Software Encryption-2001.
Nagendar Yerukala is presently working as Research Associate in CRRAO AIMSCS, Hyderabad. He did his M.Tech from NITK surathkal and M.Sc from Kakatiya university. He is pursuing his Ph.D from JNTUH Hyderabad. His areas of interest are Network security and Cryptology.

V. Kamakshi Prasad is working as Director of Evaluation and Professor in the Department of Computer Science and Engineering in JNTU Hyderabad. He obtained his PhD from the Indian Institute of Technology (IITM), Madras. He published his publications in several international journals and international conferences. He held several positions in JNTU Hyderabad. His research interests are in the areas of speech recognition, image processing, data mining and security.

Allam Appa Rao is a chairman of National Institute of Technical Teachers Training and Research, Chennai. Allam Appa Rao is a former Director of CR Rao AIMSCS, Hyderabad. He was the first to receive Ph.D from Andhra University in Computer Engineering in the year 1984. During his more than four decades of professional experience, such as first Vice Chancellor, JNTUK, Kakinada, A.P, Principal, College of Engineering (Autonomous), Andhra University. Indian Science Congress Association (ISCA) conferred him with Srinivas Ramanujan Birth Centenary Award Gold medal for his significant and lifetime contribution to the development of Science and Technology in the country specifically in the area of Computational Biology, Software Engineering and Network Security.