1. Marek LEŚNIEWICZ, 2. Piotr KOMOROWSKI, 3. Janusz ZABŁOCKI

Wojskowy Instytut Łączności – Państwowy Instytut Badawczy
ORCID: 1. 0000-0001-5505-1268; 2. 0000-0002-0752-0055; 3. 0000-0001-5429-8058

doi:10.15199/48.2023.05.13

Mikrofalowy generator losowych ciągów binarnych – konstrukcja, oprogramowanie i dowód bezpieczeństwa

Streszczenie. *W artykule przedstawiono opis elektronicznych i programistycznych rozwiązań sprzętowego, mikrofalowego generatora losowych ciągów binarnych o przepływności wyjściowej 1 Gbit/s. Szczególną uwagę poświęcono problemom programowej obsługi procesów przetwarzania ciągów o tak dużej przepływności. Opisano autorską metodę pomiarów i weryfikacji entropii generowanych ciągów oraz przedstawiono warunkowy, kryptograficzny dowód ich bezpieczeństwa.*

Abstract*. The article presents a description of electronic and programming solutions for a hardware, microwave random binary sequences generator with an output bit rate of 1 Gbps. Particular attention was paid to the problems of programmatic handling of string processing processes with such a high throughput. Have been described the method of measuring and verifying the entropy of the generated strings and presents a conditional, cryptographic proof of their security. (**Microwave random binary sequence generator – design, software and security proof**.)*

Słowa kluczowe: generacja ciągów (liczb) losowych, losowość, entropia, układy i sygnały mikrofalowe
Keywords: random sequences (number) generation, randomness, entropy, microwave circuits and signals

Wstęp

Przedmiotem pracy są elektroniczne i programistyczne rozwiązania sprzętowego, mikrofalowego generatora losowych ciągów binarnych o przepływności wyjściowej 1 Gbit/s. W poprzednim artykule [1] opisano założenia do takiej generacji, które okazały się poprawne, tzn. dały się zaimplementować na nowo zaprojektowanej platformie sprzętowej, zawierającej mikrofalowe generatory sygnałów Poissona, układ programowalny do obróbki tych sygnałów, generacji ciągu losowego i testowania jego losowości oraz układy interfejsowe pozwalające na oddawanie ciągu o tak dużej przepływności do współpracującego komputera.

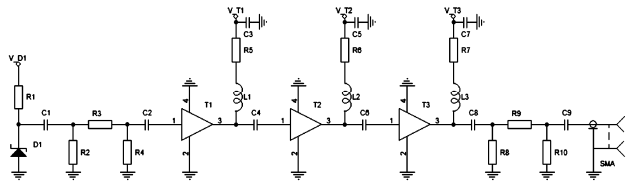
Poprawność tych założeń nie stanowiła jednak o łatwości implementacji, zwłaszcza programistycznej. Okazało się, że o ile generacja mikrofalowych sygnałów Poissona o widmie w paśmie kilku GHz oraz ich wprowadzanie i próbkowanie w układzie programowalnym nie są trudne technicznie, o tyle dalsza obróbka w czasie rzeczywistym ciągu binarnego o przepływności 1 Gbit/s stanowi duży problem, a właściwie zbiór problemów. Okazało się też, że można je rozwiązać tylko w układach programowalnych, dysponujących dużymi zasobami sprzętowymi w sensie dużej ilości elementów logicznych i dużych pamięci wewnętrznych. Wymagania te muszą być skorelowane z odpowiednią szybkością tych elementów w sensie głównie częstotliwości zegarów z pętli fazowych PLL, częstotliwości przełączania przerzutników oraz czasów zapisu i odczytu danych z wewnętrznych, statycznych pamięci S-RAM typu M10K i MLAB. W pracy tej ponownie użyto w pełni opanowanych przez autorów układów programowalnych firmy INTEL rodziny ARRIA V GT typu 5AGTMC7G3F31I3N [2]. Można dodać, że wpisuje się to w wymagania formalne, że współczesne rozwiązania układów kryptograficznych muszą być sprzętowe [3 – 6].

W dalszym ciągu zostaną przedstawione najciekawsze problemy i ich rozwiązania, które pozwoliły skonstruować właśnie sprzętowy, w pełni funkcjonalny, mikrofalowy generator losowych ciągów binarnych o przepływności wyjściowej 1 Gbit/s oraz przeprowadzić prosty dowód jego warunkowego bezpieczeństwa kryptograficznego.

Generatory mikrofalowych sygnałów Poissona

Generację sygnałów Poissona o gęstości zmian λ, przekraczającej nawet 2 GHz, przeprowadza się w układzie z rysunku 1. Zasada jego pracy jest prosta – dioda lawinowa D1, pracująca w trybie odwrotnej polaryzacji w wyniku zjawiska mikroplazmatycznego przebicia złącza p-n generuje szum binarny, modelowany sygnałem Poissona.

Odpowiednio dobrany punkt pracy tej diody w sensie jej optymalnego prądu, wymusza pożądane właściwości czasowe i widmowe sygnału Poissona. Sygnał ten jest jednak relatywnie mały, co uniemożliwia jego przetwarzanie w układach cyfrowych, zwłaszcza programowalnych. Konieczne jest więc jego wzmocnienie bez zniekształceń w funkcji czasu (zachowanie binarnego przebiegu wartości chwilowych – rysunek 3) i częstotliwości (funkcja widmowej gęstości mocy sygnału Poissona, modelowana jako $G(\omega) = G(0) / \{1 + (\omega / 4\lambda)^2\}$ – rysunek 2).

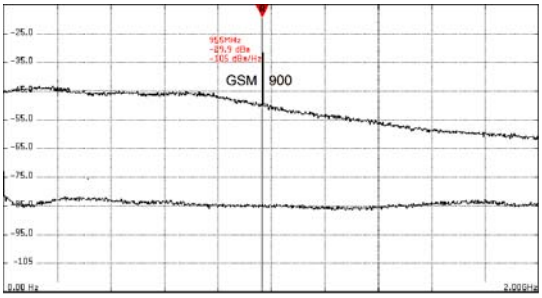


Rys.1. Układ mikrofalowego generatora sygnału Poissona

W układzie z rysunku 1 funkcję wzmacniacza liniowego pełnią dwa pierwsze układy MMIC (*Monolithic Microwave Integrated Circuit*), oznaczone jako T1 i T2. Sygnał z diody D1 ma poziom od 1 mV_{SK} do 2 mV_{SK}, więc przy typowym wzmocnieniu $K_{T1} = K_{T2} = 20$ dB na wyjściu T2 otrzymuje się poziomy od 100 mV_{SK} do 200 mV_{SK}. Poziomy te są za małe dla poprawnej pracy odbiorników transceiverów układów programowalnych, które wymagają ponadto odpowiedniego przebiegu, w konsekwencji regeneracji kształtu sygnału wejściowego. Dokonuje się ona w układzie T3, który pełni funkcje wzmacniacza ograniczającego poziomy sygnałów wyjściowych do binarnych wartości w zakresie zmienności od około +0,4 V do –0,4 V, co stanowi optymalny zakres dla wejść tych odbiorników. Ponieważ ostatni stopień układu T3 nie pracuje w trybie liniowym, to jego wzmocnienie można szacować od 6 dB do 12 dB. Wzmocnienie całego toru układów od T1 do T3, uwzględniając tłumienia układów dopasowujących R1, R2 i R3 oraz R6, R7 i R8 można więc szacować na 50 dB.

Układ o tak dużym wzmocnieniu, przetwarzający źródłowy sygnał o wartości od 1 mV_{SK} do 2 mV_{SK}, musi być odseparowany od wszelkich oddziaływań zewnętrznych w sensie zakłóceń środowiskowych i wewnątrzukładowych. Uwzględniając, że współczynnik ENR diody lawinowej wynosi około 30 dB, to dla napięcia sygnału użytecznego, tzn. sygnału Poissona rzędu 1 mV_{SK}, napięcie lokalnego szumu zakłócającego ten sygnał w samej diodzie można

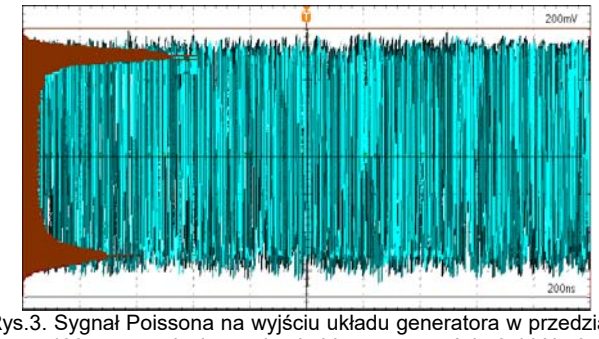
szacować na 30 μV_{SK} i nie powinien on być zwiększany przez inne, potencjalne źródła zakłóceń. Wymaga to bardzo starannego projektu płytki drukowanej, aby wyeliminować wpływ na diodę zakłóceń od strony napięcia zasilania oraz sygnałów wzmacnianych przez stopnie T1, T2 i zwłaszcza T3. Ponadto cały układ generatora sygnału Poissona musi zostać zamknięty w metalowej obudowie ekranującej o tłumienności minimum 60 dB dla częstotliwości do 18 GHz. Wyeliminuje to wszelkie zakłócenia radiowe, telewizyjne, Wi-Fi, GSM / LTE itp. Dowodem przedstawiona na rysunku 2 widmowa gęstość mocy w paśmie 2 GHz na wyjściu nieekranowanego układu. Widać na nim prążek zakłóceń pochodzący od pracującego obok telefonu komórkowego GSM, który znika pod poziomem szumów odniesienia DANL analizatora po zamknięciu obudowy ekranującej generator.



Rys.2. Widmowa gęstość mocy sygnału Poissona o λ = 1,5 GHz i szum DANL analizatora przy paśmie pomiarowym RBW = 100 kHz

Układ z rysunku 1 charakteryzuje się dobrą stabilnością w funkcji temperatury mimo, że dosyć mocno się nagrzewa, ponieważ moc wydzielana na każdym z układów T1, T2 i T3 i skojarzonych z nimi rezystorów R5, R6 i R7 sięga w sumie 1,5 W. Nie ma to jednak większego wpływu na generowane sygnały ponieważ w przedziale temperatur od 0^{O}C do 50^{O}C napięcie sygnału Poissona z diody D1 zmienia się o ±10%, a stabilność wzmocnienia układów jest rzędu 0,005 dB/^{O}C. Problemem jest jednak praca ostatniego stopnia w trybie wzmacniacza ograniczającego poziomy sygnałów wyjściowych i utrzymywanie ich równowagi na referencyjnym poziomie. Jest to niezbędne, ponieważ wynika z dwóch ważnych powodów. Pierwszego – zrównoważony sygnał Poissona pozwoli uzyskać po próbkowaniu ciąg binarny o dobrej równowadze zer i jedynek. Drugiego – opisanego w następnym punkcie.

Stabilizacja równowagi sygnału Poissona jest jednak trudnym zadaniem, ponieważ może polegać tylko na odpowiedniej korekcji punktu pracy układu T3, ale do tego celu nie może zostać użyty wyjściowy sygnał generatora, ponieważ mógłby zostać zniekształcony, czy zakłócony. Ponadto nie wiadomo, czy ustawienie nawet najlepszej równowagi w sygnałowej domenie analogowej dałoby odpowiednią równowagę w domenie cyfrowej po próbkowaniu, ponieważ mogą ją naruszyć niesymetryczne połączenie generatora z wejściem transceivera układu programowalnego, ale przede wszystkim asymetryczne właściwości wejścia próbkującego. Należy bowiem pamiętać, że pożądana równowaga zer i jedynek w wynikowym ciągu binarnym wynosi s < 0,01, a omawiane powyżej czynniki wprowadzają asymetrie rzędu ±0,05. Potwierdza to praktyka pomiarowa – nawet najlepiej ustawiona na oscyloskopie równowaga stanów +0,4 V i –0,4 V daje w ciągu binarnym nierównowagę zer i jedynek rzędu nawet 0,1. Mechanizm i układ stabilizacji równowagi zer i jedynek w wynikowym ciągu binarnym musi zatem polegać na pomiarze różnicy ich statystyk w czasie rzędu milisekund i wypracowaniu funkcji błędu, która poprzez przetwornik cyfrowo-analogowy skoryguje punkt pracy T3. Nie jest to już trudne, a zapewnia równowagę nie tylko od wpływu punktu pracy T3, a również wszystkich innych czynników – wahań napięcia zasilania, wpływu temperatury, starzenia elementów itp. Na rysunku 3 przedstawiono wynik zmiany napięcia zasilania układu o zaledwie 0,5 V, co spowodowało zachwianie równowagi stanów, zobrazowane na asymetrii rozkładu gęstości prawdopodobieństwa po lewej stronie ekranu oscyloskopu, a wynikowy ciąg binarny miał w tym stanie nierównowagę bliską s = 0,2.

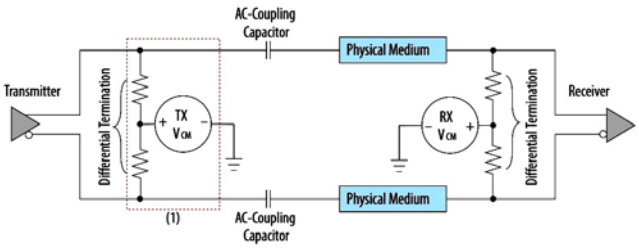


Rys.3. Sygnał Poissona na wyjściu układu generatora w przedziale czasu 400 ns – napięcia przyjmują binarne wartości +0,4 V i –0,4 V ale funkcja gęstości prawdopodobieństwa jest niesymetryczna

Układy próbkujące sygnały Poissona

Jako układów próbkujących losowe, binarne wartości sygnałów Poissona użyto odbiorników tranceiverów układów programowalnych (rys.4), pracujących w najprostszym trybie *PMA direct* [1], [2]. W tym trybie znormalizowane sygnały wejściowe są próbkowane bez żadnych dodatkowych operacji, a odczytane wartości są przepisywane do 80-bitowych buforów, skąd są odczytywane i dalej obrabiane z relatywnie wolnym zegarem o częstotliwości 1 GHz / 80 = 12,5 MHz. Koncepcyjna prostota tej operacji pociąga jednak za sobą konieczność ścisłego zachowania zasad użycia i pracy odbiorników tranceiverów, pracujących w tym trybie. Sprowadzają się one do następujących zasad:

- połączenia generatorów sygnałów Poissona z wejściami transceiverów muszą być prowadzone jako symetryczne linie paskowe o określonej geometrii ścieżek na płytce drukowanej lub dopasowanymi kablami symetrycznymi, jeśli generatory i układy programowalne znajdują się na różnych płytkach; w każdym przypadku takie połączenia powinny być jak najkrótsze i siłą rzeczy niskostratne,
- moduły i połączenia muszą być od siebie odseparowane w sensie braku wpływu przenikania przewodowego i elektromagnetycznego od innych modułów i połączeń,
- sygnał wejściowy musi być binarny o ściśle określonych poziomach; sygnały analogowe, np. szum biały o nawet większym poziomie, nie są przez nie w ogóle „widziane”,
- sygnał wejściowy musi mieć wstępnie zrównoważone poziomy binarne, które następnie w wyniku działania mechanizmu równoważenia statystyk zer i jedynek w ciągu binarnym są docelowo precyzowane w czasie rzeczywistym; sygnały o nierównomiernych rozkładach poziomów nie są poprawnie przetwarzane.



Rys.4. *Receiver* układu 5AGTMC7G3F31I3N jako układ próbkujący sygnał Poissona z generatora mikrofalowego *Transmiter* [11]

Dylematy pomiarów losowości ciągów

Testowanie ciągów losowych jest problemem mającym niezwykle obszerną literaturę [7], [8]. Autorzy przy okazji poprzednich prac mieli zaszczyt i przyjemność wymienić się opiniami na ten temat z nieżyjącym już niestety prof. Ryszardem Zielińskim. Opinię Profesora można streścić dwoma zdaniami. Pierwszym – testy statystyczne stosują jako ostateczność ci, którzy nie umieją lub nie mogą przedstawić adekwatnych modeli badanych zjawisk. Drugim – w przypadku ciągów losowych, jeżeli twórca generatora potrafi przedstawić w pełni uzasadniony, matematyczno-fizyczny model takiego generatora i obliczyć analitycznie wartość oczekiwaną oraz wariancję generowanych ciągów, to nie musi już stosować żadnych innych testów, poza właśnie sprawdzaniem ich wartości oczekiwanej i wariancji, bo inne testy są spełnione z natury rzeczy. Oczywiście opinia ta dotyczy tylko ciągów losowych w pełni i ściśle opisanych probabilistycznie, w tym przypadku jako łańcuchy Markowa 1. rzędu, z założeniem, że źródłem losowości jest zjawisko opisane procesem, tu sygnałem Poissona.

Warto w tym miejscu pokazać, że wiara w moc testów statystycznych, nawet przeznaczonych do zastosowań kryptograficznych i firmowanych przez poważne instytucje normalizacyjne, takie jak amerykański NIST [7], może być podważona logicznie i zakwestionowana doświadczalnie.

Pierwszą wątpliwością, jaką można tutaj wskazać, jest założenie, że testy te są przeznaczone do badania ciągów pseudolosowych, generowanych algorytmicznie oraz ciągów prawdziwie losowych, generowanych sprzętowo. Ciągi generowane algorytmicznie, powołując się na opinię prof. Zielińskiego, powinny mieć *a priori* udowodnione właściwości i parametry statystyczne, więc nie powinno być potrzeby badania ich testami. Tak niestety nie jest, a raczej dotąd nie było, ponieważ obecnie najnowsze konstrukcje algorytmów kryptograficznych, np. *sponge* („gąbka") mają, można powiedzieć, „genetycznie" wbudowane statystyki na referencyjnym poziomie losowości, których nie ma sensu badać testami, bo wynikają z dowodów matematycznych.

Drugą wątpliwością jest ta, że błędy losowości generatorów sprzętowych są ogólnie znane i sprowadzają się zwykle do nierównowagi statystyk zer i jedynek (*bias*) oraz korelacji niskich rzędów między dwoma, co najwyżej kilkoma sąsiednimi elementami ciągu (*correlation*). Jeśli tak, to testy do takich badań powinny być zorientowane na te właśnie rodzaje błędów, łatwo i szybko je wykrywać. Można oczywiście zakładać, że w generatorze sprzętowym ma miejsce wpływ np. 100 Hz tętnień napięcia zasilania od źle zaprojektowanego lub uszkodzonego zasilacza sieciowego. Przy generacji z przepływnością 1 Gbit/s mogłoby to skutkować anomaliami co 10 ms, a więc co 10 milionów elementów ciągu, co przy badaniu każdej próby 10 MB powinno ujawnić się dokładnie 8 razy. Znane testy nie są jednak zorientowane na szukanie takich anomalii.

Trzecia wątpliwość wynika z naturalnej właściwości testów, a mianowicie, że wykonywane są one *a posteriori* na długich próbach ciągów, a ich wynik znany jest dopiero po zapisaniu i obliczeniu parametrów statystycznych całej próby ciągu. Tymczasem w przypadku sprzętowej generacji ciągów w czasie rzeczywistym chce się mieć na bieżąco uaktualniane i oczekiwanie dobre wyniki, a jedynie w przypadku jakiś anomalii alarm o potencjalnej nielosowości.

Dalej pokazano doświadczalnie, jak łatwo jest oszukać nawet zaawansowane i dotąd uznawane testy NIST [7]. Zawierają one 15 sprawdzeń, przedstawionych w tabeli 1. Były one przedmiotem wielu pozytywnie krytycznych analiz naukowych, powstawały coraz doskonalsze wersje, wprowadzano drobne korekty [9], w końcu usunięto nawet jedno sprawdzenie, *Lempel Ziv Compression*, polegające na kompresji próby ciągu i sprawdzeniu stopnia kompresji. Było ono nie tyle naiwne – intencjonalnie nawiązywało do pojęcia *złożoności Kołmogorowa* – ile nieskuteczne nawet wobec ciągów o losowości dalekiej od doskonałości.

Tab.1. Zestaw sprawdzeń w testach NIST i ich wyniki dla ciągu doskonale losowego, ciągu *raw* z generatora bez *post-processingu* i tego ciągu pomnożonego modulo 2 przez wzór 010…101; test spełnia *P-value* > 0,01; wszystkie próby miały liczebność 10 MB

Nr	Test	losowy	*raw*	*po mod.2*
1	Frequency (Monobit)	**SUCCESS** 0.173233	FAILURE 0.000000	**SUCCESS** 0.048035
2	Block Frequency	**SUCCESS** 0.066641	FAILURE 0.000000	**SUCCESS** 0.302047
3	Runs	**SUCCESS** 0.902119	FAILURE 0.000000	FAILURE 0.000000
4	Longest Run of Ones	**SUCCESS** 0.162319	FAILURE 0.000000	FAILURE 0.000000
5	Binary Matrix Rank	**SUCCESS** 0.581598	FAILURE 0.000000	**SUCCESS** 0.588055
6	Discrete Fourier Transform	**SUCCESS** 0.525091	**SUCCESS** 0.525091	**SUCCESS** 0.525091
7	Non-overlapping Template Matching	**SUCCESS**	FAILURE / **SUCCESS**	**SUCCESS** / FAILURE
8	Overlapping Template Matching	**SUCCESS**	FAILURE	FAILURE
9	Universal Statistical (Maurer's Test)	**SUCCESS** 0.248658	**SUCCESS** 0.467363	**SUCCESS** 0.234853
10	Linear Complexity	**SUCCESS**	**SUCCESS**	**SUCCESS**
11	Serial	**SUCCESS** **SUCCESS**	FAILURE / **SUCCESS**	FAILURE / **SUCCESS**
12	Approximate Entropy	**SUCCESS** 0.711656	FAILURE 0.000000	FAILURE 0.000000
13	Cumulative Sums	**SUCCESS** 0.291775	FAILURE 0.000000	**SUCCESS** 0.050656
14	RandomExcursions	**SUCCESS**	FAILURE	**SUCCESS**
15	Random Excursions Variant	**SUCCESS** 0.848856	FAILURE 0.000000	**SUCCESS** 0.488375

Jako ciągu *raw* do testów użyto ciągu pochodzącego z generatora, charakteryzowanego przez nierównowagę zer i jedynek równą s = 0,005 i współczynnik korelacji K = 0,01. Są to typowe wartości, właściwie większości ciągów losowych generowanych przez poprawnie skonstruowane, pojedyncze generatory sprzętowe. Entropia takiego ciągu, wyznaczona z zależności $H = 1 - \{ (2s)^2 + 1/2\ K^2 \} / 2\ln 2$, wynosi 0,9999 bitu na element ciągu i w wielu publikacjach uznawana jest za doskonałą, choć wyniki 2/3 sprawdzeń z tabeli 1 są negatywne. Łatwo zauważyć, że negatywne są w tych sprawdzeniach, które są wrażliwe właśnie na błędy nierównowagi zer i jedynek oraz lokalnych korelacji.

Można teraz dokonać operacji „poprawiającej" w postaci mnożenia modulo 2 tego samego ciągu przez najprostszą, deterministyczną sekwencję 010…101, równą liczebności ciągu. Po tej operacji 2/3 wyników w trzeciej kolumnie jest już formalne pozytywne, choć entropia ciągu pozostała taka sama, mimo, że wartość s w wyniku operacji XOR została formalnie wyzerowana, a współczynnik korelacji, nie zmieniając wartości, zmienił znak na przeciwny. Entropia „*pozostała taka sama*", ponieważ wystarczy ponownie pomnożyć taki ciąg przez sekwencję 010…101, by wrócił on do poprzedniej postaci. Mogą to zrobić kryptoanalityk przeciwnika i nasz ekspert, oceniający nie tyle ciągi, ile konstrukcję generatora. Ten pierwszy nie zrobi tego jednak skutecznie, bo nie będzie miał okazji – nasz i każdy inny ekspert od razu zakwestionuje taką operację „poprawiającą". Oczywiście łatwo można sobie wyobrazić więcej operacji „poprawiających", podobnie prostych i skutecznych wobec testów i może kryptoanalityka przeciwnika, ale nie wobec eksperta, któremu należy przedstawić pełny opis koncepcji matematycznej i implementacji technicznej generatora ze schematami elektrycznymi oraz kodami oprogramowania wraz z komentarzami, szczegółowo wyjaśniającymi intencje i mechanizmy wszystkich operacji fizycznych i logicznych.

Opisanej operacji nie można zatem uznać za „ekstrakcję losowości”, czy „poprawianie” niedoskonałej losowości, a jedynie za naiwną próbę oszukania testów statystycznych. Konieczność odrzucenia takich pomysłów teoretycznie wyjaśnia nauka o informacji – żadna odwracalna operacja na ciągach nie może zmienić ich entropii, ponieważ stanowi ona ich właściwość źródłową i nie da się żadnym sposobem powiększyć, ani zmniejszyć, natomiast każda operacja nieodwracalna może ją tylko zmniejszyć [10]. Przykładem może być dodanie do siebie ciągu losowego o jednostkowej entropii i właśnie ciągu 010…101. Po takiej operacji połowa jedynek ciągu losowego zachowa swoją wartość, ale druga połowa stanie się jedynkami. Jeśli ciąg był losowy, to nie da się tego odwrócić, bo nie wiadomo, która z obecnych jedynek była poprzednio zerem. Wiadomo jednak, że statystyka jedynek zwiększyła się do 3/4, a zer spadła do 1/4, co pozwala określić entropię ciągu po takiej operacji na H = –3/4 $\log_2$ 3/4 –1/4 $\log_2$ 1/4 ≈ 0,811 bit/e < 1 bit/e, a sam ciąg można określić, jako losowy, ale już o niejednostkowej entropii. Mogłoby się wydawać, że wystarczy zestawić taki ciąg z ciągiem 010…101 i poszukać korelacji, ale będą one zerowe, niezależnie od tego, czy ich szukanie zacznie się od wzoru 101…010, czy od 010…101. Ta pozornie dziwna właściwość wynika stąd, że zmienne losowe, a więc i elementy ciągu losowego są od siebie niezależne i nie ma możliwości takiej identyfikacji. Taki sam wynik da również sumowanie dwóch ciągów losowych, ale tutaj intuicja od razu podpowie, że takich korelacji na pewno być nie może.

Widać zatem, że weryfikowanie koncepcji i jakości działania nawet modelowych wersji generatorów nie może polegać na testach statystycznych, ponieważ niesie za sobą pokusę skonstruowania generatora spełniającego siłą rzeczy ograniczony zbiór testów. Pierwszym zadaniem musi być zatem wykonanie w pełni i poprawnie opisanego oraz uzasadnionego modelu matematyczno-fizycznego mającego techniczny potencjał realizacyjny. Drugim jest wykazanie metodami analizy matematycznej, jaka będzie entropia generowanych ciągów. Trzecim – sprawdzenie metodami pomiarowymi, jaka rzeczywiście jest i czy jest zgodna.

Koncepcja pomiaru entropii w czasie rzeczywistym

Teoretyczne podstawy pomiarów entropii oczekiwanej oraz innych właściwości i parametrów ciągów losowych zostały dokładnie opisane w [11], a ich skrót w [12] i [13]. Istota takich pomiarów opiera się na odróżnieniu wprowadzonego przez Claude’a Elwooda Shannona [14] pojęcia entropii ciągu N-wymiarowych zmiennych losowych (X_1,… X_N), jako

(1) $$H(X_1,\dots X_N) = -\sum_i \mathrm{P}(X_1, \dots X_N) \log_2 \mathrm{P}(X_1,\dots X_N)$$

i entropii z próby

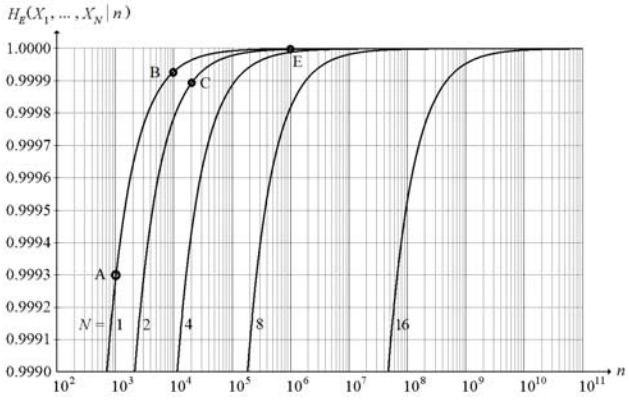
(2) $$H_p(X_1,\dots X_N) = -\sum_i \mathrm{p}(X_1, \dots X_N) \log_2 \mathrm{p}(X_1,\dots X_N) .$$

Ta pierwsza opisuje ciąg losowy *a priori* i jako taka jest niemierzalna, ponieważ operuje prawdopodobieństwami zmiennych losowych P(X_1, … X_N). Druga stanowi statystykę z próby, czyli realizacji ciągu losowego *a posteriori*, tzn. po zapisaniu próby ciągu o danej liczebności i obliczeniu średnich częstości realizacji N-elementowych sekwencji (X_1, … X_N), jako p(X_1,… X_N).

Okazuje się, entropia z próby nie jest absolutna, tzn. stała i niezależna od liczebności próby, a wykazuje taką samą zbieżność, jak wariancja próby ciągu i jest to zbieżność w relacji 1/n, gdzie n jest liczebnością próby. Dla ciągów doskonale losowych, tzn. z prawdopodobieństwami zer i jedynek równymi P(0) = P(1) = 1/2 i brakiem jakichkolwiek korelacji entropia z próby n-elementowej dla N-wymiarowej zmiennej losowej dana jest jako

(3) $$H_E(X_1, \dots, X_N \mid n) = 1 - \frac{2^{N-1}(1-1/2^N)}{n \ln 2} ,$$

co ilustruje rysunek 5.



Rys.5. Zbieżność entropii oczekiwanej ciągu doskonale losowego

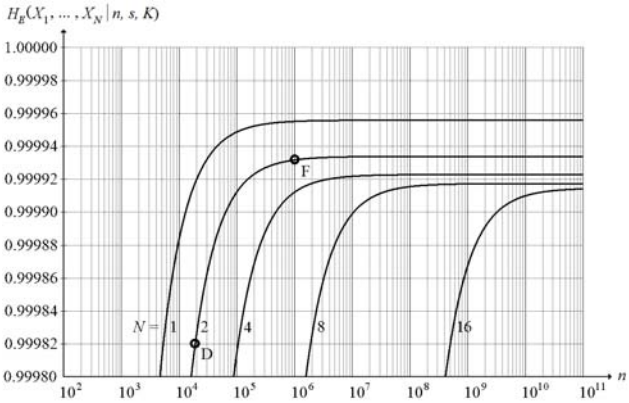
Wynika z niego, że dla N = 1, jednowymiarowej zmiennej losowej, czyli dla zer i jedynek, próba o liczebności np. n = 1.000 elementów praktycznie nigdy nie wykazuje entropii z takiej próby większej od H_R = 0,9993 (pkt.A), ale też nie jest mniejsza. Stabilizacja wartości entropii jest tym większa, im próba jest liczebniejsza, ale jest również z natury rzeczy jej wartość jest coraz większa, co wynika ze zbieżności w relacji 1/n i dla n = 10.000 osiąga już H_R = 0,99993 (B), ale również nie przekracza tej wartości, ani też nie jest od niej mniejsza. Właściwość ta została potwierdzona tysiącami prób o liczebnościach od kilobajtów do gigabajtów i żadna z nich nie wykazała, by relacja ta nie była spełniona. Relacja jest prawdziwa dla dowolnych N-wymiarowych zmiennych losowych, tzn. dla N=2, czyli par (0,0), (0,1), (1,0) i (1,0), N = 3, czyli trójek (0,0,0), (0,0,1), (0,1,0) i (0,1,1), (1,0,0), (1,0,1), (1,1,0) i (1,1,1) i wszystkich kolejnych.

W przypadku, kiedy ciąg jest modelowany jako łańcuch Markowa 1. rzędu i jest niedoskonale losowy, tzn. ma nierównowagę zer i jedynek równą P(0) – P(1) = s, a korelacje między jego sąsiednimi elementami dane są jako współczynnik korelacji K = P(0,0) + P(1,1) – P(0,1) – P(1,0), to próba o liczebności n elementów ma entropię jeszcze mniejszą, bo obniżoną dodatkowo przez wpływ s i K, zgodnie z zależnością

(4) $$H_E(X_1, \dots, X_N \mid n, s, K) =$$

$$= 1 - \frac{1}{2\ln 2}\left(\frac{2^N(1-1/2^N)}{n}(1+2K) + (2s)^2 + \frac{N-1}{N}K^2\right),$$

co dla wartości s = 0,005 i K = 0,01 ilustruje rysunek 6.



Rys.6. Zbieżność entropii oczekiwanej ciągu niedoskonale losowego

Z tych właściwości i zależności można skonstruować praktyczne kryterium identyfikacji losowości danej próby ciągu niedoskonale losowego względem hipotetycznej próby ciągu doskonale losowego. Z rysunku 5 wynika, że dla N = 2-wymiarowej zmiennej losowej próba ciągu doskonale losowego o liczebności np. n = 20.000 elementów wykazuje entropię H_{R1} = 0,9999 (C), a z rysunku 6, że próba ciągu niedoskonale losowego o takiej samej liczebności entropię H_{R2} = 0,9998 (D), czyli praktycznie taką samą, choć oczywiście nieco mniejszą, bo obniżoną o wpływ s i K. Aby lepiej rozróżnić entropie obu ciągów, należałoby zwiększyć liczebność prób i np. dla n = 1.000.000 elementów będzie odpowiednio H_{R1} = 0,999998 (E) i H_{R2} = 0,99993 (F), ale ta druga wartość już nie wzrośnie, ponieważ jest ograniczona właśnie przez s i K. O ile zatem dla prób o liczebnościach n = 20.000 elementów można by zakwalifikować drugi ciąg, jeszcze jako ciąg o entropii bliskiej entropii ciągu doskonale losowego, o tyle dla próby n = 1.000.000 elementów absolutnie nie można tego zrobić i należy go uznać za niedoskonale losowy. Na takim przykładzie można zatem skonstruować *entropijne kryterium losowości próby ciągu* – jeśli dla próby badanego ciągu o liczebności n jego entropia oczekiwana jest porównywalna z entropią hipotetycznej próby ciągu doskonale losowego o takiej samej liczebności, to nie można zaprzeczyć hipotezie, że jest to próba ciągu doskonale losowego. Kryterium można zapisać jako

$$H_E(X_1,\dots X_N \mid n, s, K) \le H_E(X_1,\dots X_N \mid n)\,, \tag{5}$$

a po podstawieniu do relacji (5) zależności (3) i (4) wynika maksymalna liczebność n próby badanego ciągu, uniemożliwiająca zaprzeczeniu takiej hipotezie, dana jako

$$n(N)_{MAX} \le \frac{2^{N/2-1}\sqrt{N\,(1-1/2^N)^3}}{(2s)^2+\dfrac{N-1}{N}K^2}\,, \tag{6}$$

co dla najważniejszej w modelu łańcucha Markowa 1. rzędu zmiennej losowej N = 2-wymiarowej sprowadza ją do prostej postaci

$$n(N{=}2)_{MAX} \le \frac{0{,}919}{(2s)^2+K^2/2}\,. \tag{7}$$

Taka metoda identyfikacji i jej kryterium mają i tę zaletę, że nie potrzebują referencyjnego wzorca hipotetycznej próby ciągu doskonale losowego – Shannon zostawił go we wzorze na entropię, jako wzorzec absolutny [14].

Można też zauważyć, że kryterium korzysta z opinii prof. Zielińskiego – w praktyce badane są wartość oczekiwana i wariancja ciągu, ponieważ zbieżność entropii oczekiwanej ma ten sam charakter i zależność, co zbieżność wariancji.

Jak zatem widać, ciągi z generatora, opisane nierównowagą zer i jedynek s = 0,005 i współczynnikiem korelacji K = 0,01, które testy NIST kwalifikowały jako prawie losowe, wedle entropijnego kryterium losowości okazują się niedoskonale losowe już dla prób ciągów o liczebnościach n > 6.000 elementów. W praktyce można i należy je znacząco, co najmniej kilka razy zwiększyć, dokonując ponadto sprawdzenia na co najmniej 3 różnych próbach ciągu, aby być pewnym niespełnienia kryterium.

Na koniec można zadać pytanie – jak długo trzeba by generować taką próbę ciągu, zakładając, że liczącą n = 50.000 elementów? Dla zakładanej przepływności 1 Gbit/s byłoby to 50 μs. Zatem, co prawda taki ciąg ma beznadziejnie słabe entropijne parametry statystyczne, ale dzięki tak skonstruowanemu kryterium można przekonać się o tym w bardzo krótkim czasie.

Post-processing* metodą *pilling-up

Nierównowaga zer i jedynek s oraz współczynnik korelacji K rzędu 0,01 są typowymi parametrami ciągów generowanych metodami sprzętowymi. Można próbować je zmniejszać poprzez dobieranie elementów elektronicznych, ale jest to kosztowne i nieefektywne, ponieważ prowadzi do co najwyżej dwukrotnego zmniejszenia typowych wartości, a więc zaledwie około czterokrotnego zwiększenia czasu identyfikacji nielosowości danej próby ciągu.

Odrzucając od razu wszelkie algorytmiczne metody „poprawiania" losowości ciągów [15], trzeba poszukać metody zmniejszającej wpływ nierównowagi zer i jedynek s oraz współczynnika korelacji K, wykorzystując jednak w dalszym ciągu entropijne kryterium losowości.

Pierwszym, który przedstawił w miarę skuteczny sposób na poprawę nierównowagi zer i jedynek w ciągu losowym był John von Neumann. Jego metoda korekcji polegała na badaniu kolejnych par elementów w ciągu i jeśli dana para była (0,0) lub (1,1), to była odrzucana, a jeśli (0,1) lub (1,0), to pobierany był tylko pierwszy element, odpowiednio z (0,1) zero, a z (1,0) jedynka. Pierwszą wadą tej metody było czterokrotne zmniejszenie liczebności ciągu wyjściowego. Drugą wadą to, że nie była to korekcja w sensie redukcji nierównowagi, a jedynie minimalizacja w relacji $2s^2$. Ale największą wadą było to, że była skuteczna w tej relacji tylko dla ciągów o elementach nieskorelowanych. Pokłosie tej metody jest takie, że do dzisiaj nikt już nie próbuje prowadzić operacji „poprawiających" na elementach tej samej próby ciągu, a jeśli próbuje to robić, to parafrazując słynną sentencję von Neumana: *is, of course, living in a state of sin*. Warto jednak zauważyć, że metoda opierała się na zastosowaniu operacji XOR i w idealnym przypadku pozwalała na zmniejszenie nierównowagi w kwadracie.

Zauważył to Mitsuru Matsui, twórca kryptoanalizy liniowej i zaproponował jako metodę *pilling-up* [15]. Polega ona na operacji XOR dwóch i więcej, ale *niezależnych* składowych ciągów losowych. Zakładając dla uproszczenia, że każdy z nich ma taką samą nierównowagę s, to po operacji XOR na M ciągach wynikowa nierównowaga wynosi $(2s)^{2M}$. W [11] dowiedziono ponadto, że jeśli ciąg charakteryzuje nie tylko nierównowaga s i ale współczynnik korelacji K, to entropia oczekiwana ciągu, stanowiącego wynik XOR dla M ciągów, dla N-wymiarowej zmiennej losowej dana jest jako

$$H_{\oplus E}(X_1,\dots,X_N \mid n, s, K) = \tag{8}$$

$$= 1-\frac{1}{2\ln 2}\left(\frac{2^N(1-1/2^N)}{n}(1+2K)+(2s)^{2M}+\frac{N-1}{N}K^{2M}\right).$$

Wynika z niej, że po operacji XOR zachowane są wszystkie właściwości entropii ciągów składowych, ale parametry, tzn. wartości nierównowagi s i korelacji K są bardzo mocno zminimalizowane w relacji $(2s)^{2M}$ i K^{2M}. Niestety, nie dotyczy to zbieżności entropii w funkcji liczebności próby i dalej jest ona dana w relacji 1/n.

Konsekwencje takiej potencjalnej operacji dla modelu ciągów z nierównowagą zer i jedynek równą s = 0,005 i współczynnikiem korelacji K = 0,01 będą następujące.

Liczebność prób wynikowych ciągów po operacji XOR na M ciągach składowych dla określenia kryterium identyfikacji losowości wynosi oczywiście

$$n_{\oplus}(N,M)_{MAX} \le \frac{2^{N/2-1}\sqrt{N\,(1-1/2^N)^3}}{(2s)^{2M}+\dfrac{N-1}{N}K^{2M}}\,. \tag{9}$$

W tabeli 2 przedstawiono wyniki obliczeń liczebności n i czasów generacji z przepływnością 1 Gbit/s dla N = 2 i różnych wartości M, czyli XOR dla M niezależnych ciągów.

Tab.2. Liczebność n próby ciągu spełniającego entropijne kryterium losowości i czas jej generacji z przepływnością 1 Gbit/s

M	n dla s = 0,005, K = 0,01	czas	komentarz
1	$6{,}13 \cdot 10^3 \approx 766$ B	6,13 μs	źle
2	$6{,}13 \cdot 10^7 \approx 7{,}66$ MB	61,3 ms	źle
3	$6{,}13 \cdot 10^{11} \approx 76{,}6$ GB	10 min.	źle
4	$6{,}13 \cdot 10^{15} \approx 766$ TB	71 dni	słabo
5	$6{,}13 \cdot 10^{19} \approx 7{,}66$ mln. x 1 TB	1940 lat	nieźle
6	$6{,}13 \cdot 10^{23} \approx 76{,}6$ mld. x 1 TB	19,4 mln lat	**wystarczy**
7	$6{,}13 \cdot 10^{27}$ = niefizyczne	194 mld lat	niefizyczne
8	$6{,}13 \cdot 10^{31}$ = niefizyczne	1940 bln lat	niefizyczne

Z przedstawionych liczb wynika, że do wartości M = 4 metoda *pilling-up* nie daje jeszcze pożądanych wyników, ale wartość M = 5 jest już akceptowalna, a wartości M ≥ 6 zadowalające. Podane czasy można porównać z czasami ochrony informacji o różnych klauzulach. *Ustawa o ochronie informacji niejawnych z 10 sierpnia 2010 roku* nie precyzuje takich czasów, ale jej poprzednie wersje i światowe odpowiedniki stanowiły, że informacje *poufne* muszą mieć gwarancję ochrony przez 5 lat, *tajne* przez 50 lat, a *ściśle tajne* bezterminowo. Pojęcie bezterminowo jest niezbyt precyzyjne, ale historyczny czas ponad 1000 lat powinien być tu adekwatny. Należy też uwzględniać poziom techniki i można sobie wyobrazić przeciwnika, dysponującego 8 mln dysków o pojemności 1 TB każdy, ale z całą pewnością na całym świecie nie ma 80 mld takich dysków, by zapisać taki ciąg do późniejszej kryptoanalizy. Wracając do pytania o wartość M, można przytoczyć anegdotę, dotyczącą Andrieja Nikołajewicza Kołmogorowa, któremu, oczywiście jeszcze jako uczniowi, oznajmiono: „*być może w twojej matematyce wystarczy tylko jeden dowód, ale my chcemy jeszcze 10 innych dowodów*". Wydaje się jednak, że zwiększanie liczby generatorów składowych ponad M = 8 byłoby niezasadne.

Wynika to również z uwarunkowań technicznych. Osiem niezależnych, odseparowanych od siebie, mikrofalowych generatorów sygnałów Poissona zmieści się na płytce drukowanej o wymiarach 120 x 120 mm^2, a typowe ilości tranceiverów w prostych i tanich układach programowalnych zawierają się od 9 do 12. Jest to adekwatna liczba: 8 wejść dla 8 generatorów i jeden transceiver do obsługi interfejsu, np. Ethernetu o przepływności nawet 10 Gbit/s. Należy też pamiętać, że założone wartości nierównowagi zer i jedynek s = 0,005 i współczynnika korelacji K = 0,01 są typowe, czyli średnie. Należy jednak uwzględnić rozrzuty i starzenie, wpływające zwłaszcza na korelacje, zależne od gęstości zmian w sygnale Poissona, wynikającej z kondycji diody lawinowej. Zawsze należy też uwzględniać klimatyczno-mechaniczne warunki pracy, zwłaszcza wpływ temperatury i wilgotności. Ponadto w procesie produkcji w technice montażu powierzchniowego praktycznie nie ma możliwości wstępnego dobierania elementów dyskretnych, ponieważ ich rozmiary są milimetrowe, więc końcówki mogłyby ulec uszkodzeniom przy próbie dotykania sondami multimetru. Zresztą lutownicze automaty montażowe wykluczają ręczne podawanie takich elementów, wyjętych z fabrycznych taśm-rolek, pobierając z nich elementy właśnie automatycznie.

Wymiana elementów, niespełniających podwyższonych wymagań, poprzez wylutowanie ze zwykle wielowarstwowej płytki drukowanej, jest bardzo ryzykowna dla trwałości padów lutowniczych i dochodzących do nich ścieżek. Dla elementów o rozmiarach od 805 do 603 (2,0x1,25 mm^2 i 1,6x0,8 mm^2) możliwy jest demontaż metodą nadmuchu gorącym powietrzem, ale elementy 402 i 201 (1,0x0,5 mm^2 i 0,6x0,3 mm^2) są praktycznie niewymienne, bo po prostu pękają przy próbie takiej termoingerencji. Oszacujmy zatem jeszcze raz liczebności n i czasy generacji z przepływnością 1 Gbit/s dla różnych wartości M, ale metodą najgorszego przypadku, dwukrotnie większej nierównowagi s = 0,01 i współczynnika korelacji K = 0,02. Wyniki ilustruje tabela 3.

Tab.3. Liczebność n próby ciągu spełniającego entropijne kryterium losowości i czas jej generacji z przepływnością 1 Gbit/s

M	n dla s = 0,01, K = 0,02	czas	komentarz
1	$1{,}53 \cdot 10^3 \approx 191$ B	1,53 μs	źle
2	$3{,}62 \cdot 10^5 \approx 4{,}53$ KB	362 μs	źle
3	$9{,}57 \cdot 10^9 \approx 1{,}12$ GB	9,57 s	źle
4	$2{,}39 \cdot 10^{13} \approx 2{,}99$ TB	6 h 39 min.	źle
5	$5{,}98 \cdot 10^{16} \approx 7{,}47$ tys. x 1 TB	1,90 lat	słabo
6	$1{,}50 \cdot 10^{20} \approx 18{,}8$ mln. x 1 TB	4740 lat	nieźle
7	$3{,}74 \cdot 10^{23} \approx 46{,}8$ mld. x 1 TB	11,9 mln lat	**wystarczy**
8	$9{,}35 \cdot 10^{26}$ = niefizyczne	29,6 mld lat	niefizyczne

Tym razem z przedstawionych liczb wynika, iż dopiero dla M = 6 można by uznać, że taki generator będzie się nadawał do zastosowań o klauzuli *ściśle tajne*. Precedens Kołmogorowa zdaje się jednak wskazywać, że lepiej od razu przyjąć liczbę M = 8, co jak już opisano, stanowi nie tylko o zadowalającym bezpieczeństwie, ale jest w miarę łatwo realizowalne technicznie. Widać jednak, że tylko dwukrotne obniżenie wymagań na nierównowagę s i współczynnik korelacji K spowodowało bardzo istotną zmianę wyników kryteriów w funkcji liczby generatorów M, co wynika oczywiście z „mocy" potęgowej funkcji x^{2M}.

Powyższa analiza może zostać uznana za *warunkowy dowód bezpieczeństwa generowanych ciągów*. Warunkowy, ponieważ abstrahując od czasu 29,6 mld lat, stanowiącego ponad dwukrotny czas istnienia Wszechświata, nie stanowi on czasu *bezterminowego*. Można jednak zadać pytanie, co w może dać kryptoanalitykowi przeciwnika wiedza, okazywana w powszechnie dostępnej publikacji? Gdyby nawet zdobył taki generator lub wykonał jego wierną kopię, aby potwierdzić jego właściwości i parametry, to dowie się tego samego, więc znacznie prościej i nieporównanie taniej będzie mu po prostu przeczytać ten artykuł i nie tracić czasu.

Można zatem powiedzieć, że zastosowanie ciągów z tego generatora, przy zachowaniu ich sekretu, daje w praktyce całkowitą gwarancję bezpieczeństwa. Nie zmieni tego perspektywa nadciągających komputerów kwantowych i póki ktoś nie obali teorii Shannona, można być tego pewnym.

Tab.4. Zestaw sprawdzeń w testach NIST i ich wyniki dla ciągu doskonale losowego, ciągu *raw* z generatora bez *post-processingu* i dwóch różnych ciągów *raw* pomnożonych modulo 2 (*pilling-up*); test spełnia *P-value* > 0,01; wszystkie próby miały liczebność 10 MB

Nr	Test	losowy	*raw 1 lub 2*	*XOR 1 i 2*
1	Frequency (Monobit)	**SUCCESS** 0.127732	FAILURE 0.000000	**SUCCESS** 0.319277
2	Block Frequency	**SUCCESS** 0.352995	FAILURE 0.000000	**SUCCESS** 0.510156
3	Runs	**SUCCESS** 0.083938	FAILURE 0.000000	**SUCCESS** 0.040833
4	Longest Run of Ones	**SUCCESS** 0.162319	FAILURE 0.000000	**SUCCESS** 0.395526
5	Binary Matrix Rank	**SUCCESS** 0.581598	FAILURE 0.000000	**SUCCESS** 0.693266
6	Discrete Fourier Transform	**SUCCESS** 0.525091	**SUCCESS** 0.525091	**SUCCESS** 0.436751
7	Non-overlapping Template Matching	**SUCCESS**	FAILURE / **SUCCESS**	**SUCCESS**
8	Overlapping Template Matching	**SUCCESS**	FAILURE	**SUCCESS**
9	Universal Statistical (Maurer's Test)	**SUCCESS** 0.248658	**SUCCESS** 0.467363	**SUCCESS** 0.652470
10	Linear Complexity	**SUCCESS**	**SUCCESS**	**SUCCESS**
11	Serial	**SUCCESS** **SUCCESS**	FAILURE / **SUCCESS**	**SUCCESS** **SUCCESS**
12	Approximate Entropy	**SUCCESS** 0.711656	FAILURE 0.000000	**SUCCESS** 0.504092
13	Cumulative Sums	**SUCCESS** 0.240763	FAILURE 0.000000	**SUCCESS** 0.323173
14	RandomExcursions	**SUCCESS**	FAILURE	**SUCCESS**
15	Random Excursions Variant	**SUCCESS** 0.848856	FAILURE 0.000000	**SUCCESS** 0.113990

Dla potwierdzenia skuteczności metody *pilling-up* można wziąć tylko M = 2 ciągi o nierównowadze równej s = 0,005 i współczynniku korelacji K = 0,01 i poddać je operacji XOR. Dla porównania badanie ponownie wykonano zestawem testów NIST, a jego wyniki zawiera tabela 4.

Widać, że z wyniku operacji XOR na zaledwie dwóch ciągach, ciąg wynikowy wg testów NIST jest już doskonale losowy. Niestety, okazywane miary 0,9 > *P-value* > 0,01 nic nie mówią o poziomie losowości i różne próby ciągu nawet doskonale losowego mogą otrzymać skrajne oceny, mimo, że ciąg ten ma udowodnioną entropię $1 > H > 1 - 10^{-32}$.

Z drugiego wiersza tabeli 2 wynika jednak, że kryterium entropijne zdyskwalifikowałoby taką niedoskonale losową próbę w czasie rzędu 100 ms, oczywiście pod warunkiem, że zostałaby ona dostarczona z przepływnością 1 Gbit/s.

Podsumowując próby zastosowania testów NIST do badania losowości ciągów można powiedzieć, że:

- umożliwiają one odróżnianie prób ciągów ewidentnie nielosowych od losowych,
- odróżnianie to oparte jest na mierze nierównowagi zer i jedynek; jeśli próba ma niewielką nierównowagę, to ma szansę spełnić większość innych sprawdzeń,
- intencjonalna deklaracja twórców testów, że minimalne liczebności prób w tych testach powinny przekraczać 100, czy 1000 elementów wskazuje, że nie były one konstruowane dla badań ciągów o liczebnościach rzędu 1 GB, a dopiero próby o takich liczebnościach umożliwiają wiarygodną ocenę stacjonarności i ergodyczności, a więc stabilności stochastycznej źródeł entropii i generowanych na ich postawie ciągów, co wynika z zależności (6) i (9),
- wartości *P-value* nic nie mówią o poziomie losowości całej próby; ta sama próba ciągu losowego testowana np. w przedziałach 9 MB i 10 MB zawsze otrzymuje kwalifikację *SUCCESS*, ale nigdy z tymi samymi wartościami; można odnieść wrażenie, że oddawana przez program wartość *P-value* nie dotyczy oceny całej próby, a stanowi tylko bieżącą wartość obliczoną dla ostatniego odcinka ciągu.

Na koniec tych analiz można sprawdzić, co przynoszą one w sensie praktycznych pomiarów, np. nierównowagi zer i jedynek s. Dla s = 0,005 interpretacja jest prosta – w sygnale Poissona o poziomach napięć ±0,4 V wartość średnia będzie większa o 20 mV, co można zmierzyć nawet multimetrem. Ale jaka będzie ekwiwalentna różnica po operacji XOR na M = 4 ciągach? Oczywiście $s = 10^{-16}$, co już trudniej zinterpretować, bo najlepszy na świecie wzorzec częstotliwości, czyli fontanna cezowa F2, posiadana przez amerykański NIST, ma dokładność właśnie 10^{-16}. Gdyby chcieć zobrazować tę liczbę, to odpowiada ona dokładności pomiaru odległości Ziemi od Słońca z błędem około 10 μm, czy czasu istnienia Wszechświata z dokładnością 1 minuty. Dla M = 8 i $s = 10^{-32}$ trudno wskazać jakieś realne przykłady.

Założenia do pomiarów entropii w czasie rzeczywistym

Po tym dość długim, ale niezbędnym wyjaśnieniu istoty dowodu bezpieczeństwa kryptograficznego, wynikającego z analiz, wskazujących na konieczności pomiarów entropii, można przystąpić do opisu realizacji tych pomiarów.

Jeśli zostanie do tego wykorzystany 80-bitowy bufor, odbierający dane z przepływnością 1 Gbit/s z odbiornika transceivera układu programowalnego, to dalsze operacje będą już proste w sensie zegara bufora o częstotliwości 1 GHz / 80 = 12,5 MHz, ale w sensie organizacji 80-bitowej dalej nie będą się już odbywały w czasie rzeczywistym, ale w odpowiednio zorganizowanych automatach logiczno-pamięciowych. Uwzględniając zasadę oddzielnego badania nierównowagi zer i jedynek s oraz współczynnika korelacji K, należy jednak badać je dla tej samej próby ciągu, tzn. te same dane przesyłać do automatów pomiaru nierównowagi i korelacji. Należy zatem wybrać jedną liczebność próby n, odpowiednio dużą, aby identyfikacja była dokładna, a więc jednoznaczna. Dla wartości s i K rzędu 0,01 taka liczebność powinna wynosić zatem nie minimalne n = 1500 elementów, ale co najmniej kilka razy więcej, jednak dla dużej dokładności pomiaru warto ją zwiększyć do np. n = 64.000 elementów. Pozwoli to na bardzo precyzyjne pomiary współczynnika korelacji K i możliwość śledzenia stanu diody lawinowej. Dioda ta „wypala" się w czasie, co skutkuje zmniejszaniem gęstości zmian λ. Ponieważ $K = \exp(-2\lambda / f_P)$, to znając dokładną wartość K można na bieżąco obliczać λ i łatwo kontrolować stan diody, a nawet prognozować jej żywotność. Liczebność n = 64.000 elementów nie jest duża w sensie czasu generacji, bo dla 1 Gbit/s odpowiada zaledwie 64 μs. Jest to czas, w którym trudno spodziewać się większych zmian warunków pracy całego generatora, zatem taki ciąg badań można uznać za ciągły proces kontroli jego stanu w czasie rzeczywistym.

W artykule [1] opisano koncepcję bezpośredniego pomiaru nierównowagi zer i jedynek metodą zliczania ich wystąpień w próbie n-elementowej i obliczania $s = (n_0 - n_1) / N$, gdzie $n_0 + n_1 = N$, lub prościej $s = \pm|1/2 - n_0 / 2N| = \pm(-|1/2 - n_1 / 2N|)$, a znak s będzie zależał od przewagi liczebności jednej z nich. Nawet przy minimalnej wartości s jej znak jest ważny, ponieważ wartość ta będzie przetwarzana na sygnał błędu, sterujący punktami pracy układu generatora sygnału Poissona w celu stabilizacji równowagi zer i jedynek w wynikowym ciągu binarnym. Mechanizm będzie musiał zatem działać na zasadzie bieżącego obliczenia wartości cyfrowej s i podania jej na wejście przetwornika cyfrowo-analogowego, a napięcie z jego wyjścia, jako funkcja błędu, skoryguje punkt pracy T3. Każda taka operacja będzie powtarzana co 64 μs, a w takim czasie punkty pracy układu nie zmienią zbytnio swoich parametrów.

W artykule opisano też koncepcję pośredniego pomiaru współczynnika korelacji K, jako wyniku pomiaru średniej liczby zmian z zer na jedynki, która w próbie ciągu doskonale losowym wynosi 1/4 liczebności próby, a w niedoskonale losowym jest pomniejszona o wartość współczynnika korelacji K w relacji $1/4\ (1 - (2s)^2 - K)$, praktycznie 1/4 (1 – K). Mechanizm pomiaru powinien więc działać na zasadzie pomiaru średniej częstości takich zmian w odpowiednio licznej próbie ciągu, oczywiście takiej samej, jak do pomiaru nierównowagi zer i jedynek.

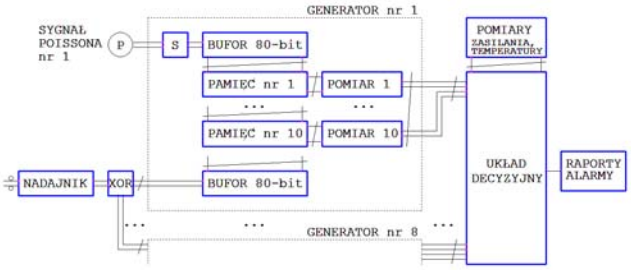
Jak już jednak wspomniano, w zastosowanych układach programowalnych firmy INTEL rodziny ARRIA V GT typu 5AGTMC7G3F31I3N, poza samym próbkowaniem, dalsze operacje z zegarem o częstotliwości 1 GHz są niemożliwe. Mimo że maksymalne częstotliwości zegarów dla funktorów logicznych i przerzutników oraz komórek pamięci sięgają 600 MHz, to zbudowane z nich, bardziej rozbudowane automaty logiczno-pamięciowe, mogą pracować z zegarami o częstotliwościach sięgających 120÷160 MHz. Praktycznie, uwzględniając niezbędne marginesy bezpieczeństwa, zasadniczo na przewidywany zakres temperatur pracy, tylko nieznacznie mogą one przekraczać 100 MHz. Jeśli jednak uwzględnić, że automaty te operują na słowach 80-bitowych, to wynika stąd możliwość pracy z ekwiwalentną przepływnością 80 bit x 100 MHz = 8 Gbit/s, a więc znacznie większą od 1 Gbit/s. Problemem inżyniera-programisty jest więc taki projekt i konfiguracja układu programowalnego, aby ten potencjał skutecznie wykorzystać.

Naturalnym, choć niełatwo realizowalnym mechanizmem przetwarzania takiego strumienia danych jest jego podział na odpowiednio mniejsze części i przetwarzanie potokowe. W przypadku ciągów losowych jest to o tyle proste, że poza elementową, nie mają one żadnej blokowej struktury i można na nich operować nawet ze stratami, ponieważ usunięcie dowolnego fragmentu takiego ciągu nie zmienia jego losowości w sensie ani właściwości, ani parametrów.

Nie może to oczywiście oznaczać żadnej intencjonalnej niedbałości w projektach programistycznych. Stanowi jednak pewną właściwość, którą można wykorzystać w przypadku niesynchronicznego przetwarzania potokowego. Jeśli np. interfejsem wyjściowym będzie Ethernet 1000Base-T, to formatowanie i odbiór przenoszonych przez niego ramek mogą być wstrzymywane przez protokół jego obsługi i czekające na to fragmenty ciągu mogą być odrzucane, bo w kolejce są gotowe następne. Taka sytuacja może dotyczyć komunikacji z potwierdzaniem poprawnego odbioru, np. typu TCP. Natomiast w przypadku komunikacji typu UDP układ odbiorczy może nie nadążać z odbiorem i niektóre pakiety mogą być tracone. Jak jednak powiedziano, nie ma to znaczenia dla zachowania losowości odbieranych ciągów a jedynie dla sprawności ich przesyłania. W każdym takim przypadku warto jest jednak zastosować jakiś wskaźnik tej sprawności, np. numerować próby ciągu i dać odbiorcy możliwość śledzenia systematyczności nadawania i odbioru. Taki śledzący mechanizm nadawczy powinien działać również po stronie nadawcy, choćby w celu wykluczenia możliwości powtórnego nadania tej samej próby ciągu, co w przypadku ciągów losowych stanowiłoby błąd kardynalny.

Układ do pomiarów entropii w czasie rzeczywistym

Schemat blokowy generatora, realizującego potokowe pomiary parametrów s i K, ilustruje rysunek 7.



Rys.7. Schemat blokowy generatora; generatory 2÷8 są takie same

Zasady pracy układu są następujące:

- dane z 80-bitowego bufora układu próbkującego S są kolejno, sekwencyjnie przepisywane do dziesięciu 64 kilobitowych partycji wewnętrznej pamięci S-RAM,
- każda pamięć ma własny zestaw automatów testujących wartości s i K z zegarem o częstotliwości 100 MHz,
- każdy automat testujący po obliczeniu wartości s i K przesyła je w postaci raportu do automatu decyzyjnego,
- ponieważ ciągi losowe podlegają prawu wielkich liczb, to próby o skończonej liczebności mają parametry zawarte w przyjętych przedziałach ufności, które jednak mogą być incydentalnie przekraczane, jako tzw. lokalne anomalie,
- automat decyzyjny rozstrzyga, czy dana anomalia jest krytyczna; w danej próbie dopuszcza się wystąpienie dwóch kolejnych anomalii po sobie, ale trzecia anomalia stanowi błąd sprzętowy i powód do alarmu,
- automat decyzyjny otrzymuje takie raporty od wszystkich M = 8 generatorów; w danym odcinku czasu dopuszcza się wystąpienie równoległych anomalii w dwóch generatorach składowych; anomalia w trzecim generatorze stanowi błąd sprzętowy i powód do alarmu,
- powrót wszystkich generatorów do pracy w przyjętych przedziałach ufności zeruje wszystkie liczniki anomalii,
- próby z 8 generatorów, spełniające powyższe kryteria (bez eliminacji wykazujących anomalie) są przepisywane do 80-bitowego bufora, dalej poddawane równoległej operacji XOR i nadaniu do układu interfejsu wyjściowego,
- do współpracującego komputera wraz z nadawanymi próbami ciągów, ale oddzielnym interfejsem, wysyłane są raporty o parametrach s, K, λ, zasilaniu, temperaturze itp.

Rzeczywisty generator MGCL-1G

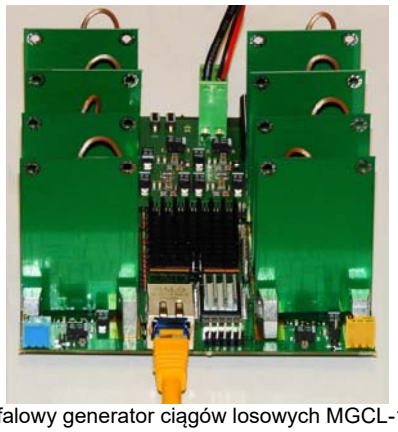
Przedstawione powyżej opisy i analizy miały charakter ilustracyjny. W dalszym ciągu zostanie opisana rzeczywista konstrukcja i osiągi opracowanego w WIŁ-PIB modelu generatora o symbolu MGCL-1G (mikrofalowy generator ciągów losowych o przepływności wyjściowej 1 Gbit/s).

W analizach wielokrotnie podnoszony był problem kompatybilności elektromagnetycznej generatora z punktu widzenia jego poprawnej pracy w zakłóconym środowisku elektromagnetycznym, ale jednocześnie braku jego wpływu na zakłócanie tego środowiska. W tym przypadku problem jest szczególnie krytyczny, ponieważ:

- jako źródła szumów mikrofalowych używane są diody, generujące ultra szerokopasmowe sygnały o relatywnie bardzo małym poziomie napięć rzędu 1 mV_{SK}, które mogą być zewnętrznie zakłócane przez promieniujące urządzenia radiowe, telewizyjne, Wi-Fi, GSM / LTE itp.
- układ programowalny sam w sobie stanowi źródło silnych zakłóceń promieniowanych i przewodzonych,
- generatory składowe mogą zakłócać się wzajemnie drogą promieniowania i przewodzenia, co spowoduje, że ich sygnały wyjściowe będą skorelowane, generowane z nich ciągi losowe nie będą niezależne, a w takim przypadku wnioski i zależności dotyczące ciągów wynikowych po operacji XOR będą nieprawdziwe,
- jakiekolwiek promieniowanie czy przewodzenie sygnałów z dowolnego elementu generatora do zewnętrznego środowiska elektromagnetycznego może stanowić jego zakłócenie, ale przede wszystkim zagrożenie emisją ujawniającą, która dla zastosowań kryptograficznych musi zostać wykluczona.

W wyniku makietowania, sprawdzeń i pomiarów różnych opcji układu okazało się, że optymalną stanowi konstrukcja blokowa ośmiu odseparowanych od siebie, mikrofalowych generatorów składowych, połączonych kablowo z układem programowalnym 5AGTMC7G3F31I3N, wydzielony moduł zasilania oraz interfejsy Ethernet 1000Base-T i USB 2.0 HS, przedstawione na rysunku 8. Taka konfiguracja zapewnia:

- brak identyfikowalnych zakłóceń i przeników sygnałów między poszczególnymi blokami generatora,
- łatwość serwisowania w sensie demontażu i wymiany generatorów składowych,
- znacząco zredukowane rozmiary oraz prostotę projektu, wykonania i montażu 6-warstwowej płytki drukowanej,
- możliwość łatwego zekranowania elektromagnetycznego całości konstrukcji szczelną obudową metalową,
- pełna zgodność z układem referencyjnego generatora ciągów losowych, opisanego w dokumencie ITU-T [4].



Rys.8. Mikrofalowy generator ciągów losowych MGCL-1G

Zgodność tę można opisać następująco:

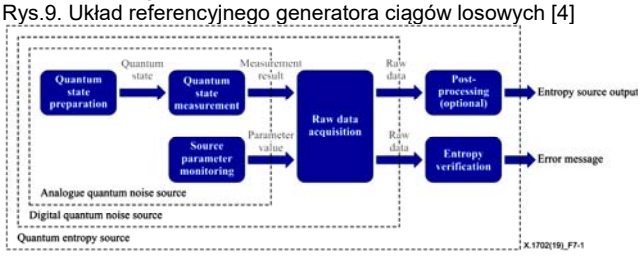
- *analogue quantum noise source* to generatory sygnałów Poissona, pochodzących ze wzmocnienia źródłowych szumów diod mikrofalowych (*quantum state preparation*),

dających na wyjściach binarne sygnały Poissona (*quantum state*),

- *quantum state measurement* to układy próbkowania aktualnych stanów, czyli chwilowych wartości sygnałów Poissona, odbieranych i zapisywanych przez układ programowalny,jako ciągi binarne (*measurement results*),
- *source parameter monitoring* to układy monitorowania źródeł, domyślnie pomiarów nierównowagi i korelacji w każdym z ciągów składowych (*parametr value*), a więc entropii ciągów składowych,
- *raw data acquisition*, czyli układ akwizycji zawiera się w układzie monitorowania,
- generatory składowe z układami próbkowania i monitorowania stanowią *digital quantum noise source*, czyli cyfrowe źródła ciągów (*raw data*) o entropii teoretycznie określonej *a priori* i na bieżąco weryfikowanej, tzn. technicznie mierzonej *a posteriori* (*entropy verification*),
- w przypadku niespełnienia przez ciągi składowe zakładanych wartości entropii (*entropy verification*), następuje wstrzymanie generacji i wysłanie komunikatu o błędnym działaniu (*error message*), tutaj poprzez interfejs USB 2.0 HS,
- jako układ *post-processing* można uznać operację XOR,
- jeśli próby ciągów składowych spełniają kryteria losowości w sensie zakładanej entropii, to wynikowy ciąg po operacj XOR jest oddawany na wyjście operacyjne Ethernet 1000Base-T (*entropy source output*) generatora ciągów losowych, czyli *quantum entropy source*,
- weryfikacja entropii polega na badaniu, tzn. pomiarach entropii ciągów składowych, a nie ciągu wynikowego, stąd przy założeniu spełnienia wymagań przez ciągi składowe ciąg wynikowy ma zawsze referencyjną, praktycznie jednostkową entropię.

Na koniec kilka słów o rzeczywistych właściwościach i parametrach generatora.

Rys.9. Układ referencyjnego generatora ciągów losowych [4]



Rys.9. Układ referencyjnego generatora ciągów losowych [4]

Generator nadaje wytwarzane ciągi *on line*, w pakietach ethernetowych o polach danych długości 1500 oktetów, bez fragmentacji. Pozwala to, po odrzuceniu oktetów ze wzorem synchronizacji, adresami sieci, sumami kontrolnymi itp., osiągnąć praktycznie 970 Mbit/s. W praktyce sprawność transferu w największym stopniu zależy od zdolności odbioru i zapisu ciągów na dyski przez współpracujący komputer. Stosując optymalizowane oprogramowanie i dyski SSD w praktyce osiąga się zapis z przepływnością 120 MB/s, co pozwala zapisać w ciągu doby dysk lub macierz dysków SSD o pojemności co najmniej 10 TB.

Pobór mocy przez generator jest znaczący i wynosi około 25 W. Ponieważ generator musi być zamknięty w szczelnej obudowie, niezbędne jest chłodzenie przez wymuszony wiatrakiem, wewnętrzny obieg powietrza i odprowadzanie ciepła właśnie obudową, stanowiącą radiator konwekcyjny. Mimo to temperatura na najcieplejszych elementach sięga 50^{O}C, co nie stanowi jednak problemu eksploatacyjnego, ponieważ generator nie będzie urządzeniem polowym.

Podsumowanie

Opisane elektroniczne i programistyczne rozwiązania sprzętowego, mikrofalowego generatora binarnych ciągów losowych o przepływności wyjściowej 1 Gbit/s dowiodły w działaniu poprawności analiz i koncepcji tak unikatowego urządzenia, nieznanego dotąd z jawnej literatury. Opisy można uznać za kompletne, ponieważ w pełni wyjaśniają zasady i wyniki działania generatora, który można uznać za źródło prawdziwie i doskonale losowych ciągów binarnych.

Technologie, zastosowane do konstrukcji nie są proste i tanie, ale przy założonym celu nie dałoby się go osiągnąć prostszymi metodami i tańszymi środkami, zgodnie z tezą, że: *wszystko powinno być tak proste, jak to tylko możliwe, ale nie prostsze*.

Uzyskane osiągi w postaci przepływności 1 Gbit/s nie predestynują generatora do jakiś lokalnych aplikacji, ale raczej źródłowego elementu ogólnodostępnego serwera ciągów losowych do zastosowań naukowo-technicznych itp.

Przedstawiony dowód bezpieczeństwa kryptograficznego generowanych ciągów może być przykładem unikatowego połączenia i wykorzystania teorii procesów stochastycznych oraz nauki o informacji. Można zauważyć, że przesłanki do takiego podejścia i koncepcji rozwiązania technicznego istniały już kilkadziesiąt lat temu, ale umożliwiły je dopiero współczesne technologie mikrofalowe i mikroelektroniczne.

Na stronie internetowej Wojskowego Instytutu Łączności znajdują się rzeczywiste próby ciągów o różnej liczebności, które mogą zostać użyte do analiz i innych zastosowań. Autorzy będą bardzo wdzięczni za opinie i uwagi dotyczące przedstawionej teorii generacji oraz rozwiązań technicznych.

***Autorzy**: dr hab. inż. Marek Leśniewicz, profesor WIŁ-PIB, mgr inż. Piotr Komorowski, Janusz Zabłocki, Zakład Kryptologii WIŁ-PIB, 05-130 Zegrze, Warszawska 22A, E-mail: m.lesniewicz@wil.waw.pl*

LITERATURA

[1] Leśniewicz M., Komorowski P., Zabłocki J., Analiza możliwości sprzętowej generacji losowych ciągów binarnych z przepływnością 1Gbit/s, *Przegląd Elektrotechniczny*, 2023, nr. 4j
[2] Arria V Device Handbook, Volume 1: Device Interfaces and Integration, Volume 2: Transceivers, *INTEL (ALTERA)*, 2020.
[3] AC/322-D - Infosec Technical and Implementation Guidance on Cryptographic Mechanisms in Hardware and Software, *NATO*.
[4] Recommendation X.1702, Quantum noise random number generator architecture, *ITU-T*, 11/2019.
[5] Barker E., Kelsey J., Recommendation for Random Bit Generator (RBG) Constructions, *NIST Special Publication 800-90C*, 2016.
[6] Turan M.S. et al: Recommendation for the Entropy Sources Used for Random Bit Generation, NIST Special Publication 800-90B, 2018.
[7] Rukhin A., et al.: A Statistical Test Suite for Random Number Generators and Pseudorandom Number Generators for Cryptographic Applications, *NIST Special Publication 800-22 rev.1a*, 2010 (nierekomendowane przez NIST od 19.04.2022 r).
[8] Wieczorkowski R., Zieliński R., Komputerowe generatory liczb losowych, *WNT*, 1997.
[9] Kim S.J., Umeno K., Hasegawa A., Corrections of the NIST Statistical Test Suite for Randomness, *arXiv:nlin/0401040*
[10] Seidler J., Nauka o informacji, *WNT*, 1983.
[11] Leśniewicz M., Sprzętowa generacja losowych ciągów binarnych, *Wydawnictwo Wojskowej Akademii Technicznej*, 2009.
[12] Leśniewicz M., Expected Entropy as a Measure and Criterion of Randomness of Binary Sequences. *Przegląd Elektrotechniczny*, 2014, nr 1.
[13] Leśniewicz M., Analyses and Measurements of Hardware Generated Random Binary Sequences Modeled as Markov Chains. *Przegląd Elektrotechniczny*, 2016, nr 11.
[14] Shannon C.E., Communication Theory of Secrecy Systems, *Bell System Technical Journal*, vol.28, 1949.
[15] Kwok SH., et all: A Comparison of Post-Processing Techniques for Biased Random Number Generators, *Proceedings of the 5th IFIP WG 11.2 International Conference on Information Security Theory and Practice*, 2011.