Tính độc lập giữa các phép kiểm tra thống kê trong NIST SP 800-22

Đỗ Đại Chí, Hoàng Đình Linh, Trương Minh Phương

***Tóm tắt*—Trong bài báo này, chúng tôi nghiên cứu về sự tương quan giữa các phép kiểm tra thống kê của NIST SP 800-22. Cụ thể, dựa vào một số phương pháp thống kê lý thuyết đã biết [3], [4], [7], [8], [11] (bao gồm: *hệ số tương quan Pearson*, *tỷ số FF*, *thông tin tương hỗ*, *độ bao phủ và độ hiệu quả bao phủ*), chúng tôi đã phân tích tính tương quan giữa các phép kiểm tra thống kê của NIST SP 800-22 thông qua đánh giá từ kết quả thực nghiệm. Từ kết quả thực nghiệm, chúng tôi đã phát hiện thêm mối tương quan mới giữa một số kiểm tra thống kê trong NIST SP 800-22 trong ngữ cảnh so sánh với các kết quả nghiên cứu đã biết.**

***Abstract*—In this paper, we analysis of the correlation between the statistical tests of NIST SP 800-22 that used as a standard in the evaluation of randomness. More precisely, based on several known theoretical statistical methods in [3], [4], [7], [8], [11] (including: *Pearson correlation coefficient*, *FF ratio*, *mutual information*, *coverage and effectiveness of coverage*), we analyzed the correlation between statistical tests of NIST SP 800-22 through evaluation of experimental results. Furthermore, our experiment results which are compared to known results show the existence of statistical dependencies between the tests in the NIST SP 800-22 that have not been previously detected.**

***Keywords*—Pearson's correlation coefficient, mutual information, statistical randomness tests, coverage.**

***Từ khoá*—Hệ số tương quan Pearson, thông tin tương hỗ, các phép kiểm tra thống kê về tính ngẫu nhiên, tính độc lập, độ bao phủ.**

I. Giới thiệu

Các số ngẫu nhiên đóng vai trò rất quan trọng trong mật mã học [9]. Để tạo ra các số ngẫu nhiên chất lượng cao (có độ bất định cao), các bộ tạo số ngẫu nhiên thực sự dựa trên các quá trình vật lý thường được sử dụng. Tuy nhiên, trong nhiều ứng dung mật mã việc sử dụng các bộ tạo số ngẫu nhiên thực sự lại tỏ ra kém hiệu quả bởi chi phí cao hoặc cần các phần cứng chuyên dụng. Do đó, các hệ thống này thường ưu tiên sử dụng *bộ sinh số giả ngẫu nhiên* (PRNG) dựa trên các thuật toán tất định. Trong mật mã, PRNG có thể được sử dụng để tạo ra các thành phần ngẫu nhiên an toàn cho các giao thức mật mã như: *giao thức xác thực*, *lược đồ chữ ký số*, *giao thức không tiết lộ tri thức*, *việc sinh khóa mã hóa*,...

Việc sử dụng các bộ sinh số ngẫu nhiên yếu có thể dẫn đến khả năng kẻ tấn công phá vỡ toàn bộ hệ thống mật mã. Do đó, chất lượng của bộ sinh số giả ngẫu nhiên PRNG là một khía cạnh rất quan trọng đối với độ an toàn của một lược đồ mật mã, cần được nghiên cứu và xác định một cách rất cẩn thận. Vì cách tiếp cận lý thuyết để chỉ ra bằng chứng về chất lượng tính ngẫu nhiên của PRNG là không khả thi, nên ta thường sử dụng cách tiếp cận thống kê dựa vào việc quan sát các dãy mẫu được tạo bởi bộ sinh và so sánh với các dãy ngẫu nhiên để đưa ra đánh giá. Bằng cách xem xét các tính chất của dãy ngẫu nhiên nhị phân, thì các phép kiểm tra thống kê khác nhau có thể được thiết kế nhằm đánh giá khẳng định rằng "*một dãy được sinh bởi một nguồn ngẫu nhiên hoàn hảo*".

Đã có nhiều bộ kiểm tra khác nhau được đề xuất như NIST SP 800-22, Diehard, Dieharder, ENT,... trong đó NISTS SP 800-22 là bộ kiểm tra được sử dụng phổ biến nhất để đánh giá tính ngẫu nhiên của bộ sinh số PRNG.

Trong bài báo này, chúng tôi trình bày một cách đầy đủ và chi tiết về việc đánh giá mối tương quan giữa các kiểm tra trong bộ kiểm tra NIST SPP 800-22 này.

Các kết quả nghiên cứu liên quan: Tính độc lập trong các kiểm tra thống kê của NIST SP800-22 đã được phân tích bởi một số nhà nghiên cứu. Soto [10] đã thảo luận về nhu cầu phân tích tính độc lập và chỉ ra rằng "các phép kiểm tra thống kê (trong bộ test thống kê) nên *độc lập* để đạt được kết quả đáng tin cậy". Soto cũng đề cập đến vấn đề *độ bao phủ kiểm tra*: các phép kiểm tra thống kê có thể độc lập, nhưng thực tế một hệ thống có thể không sử dụng hết tất cả các phép kiểm tra thống kê trong việc đánh giá tính ngẫu nhiên.

Bài báo được nhận ngày 23/02/2021. Bài báo được nhận xét bởi phản biện thứ nhất ngày 25/3/2021 và được chấp nhận đăng ngày 31/3/2021. Bài báo được nhận xét bởi phản biện thứ hai ngày 29/3/2021 và được chấp nhận đăng ngày 01/4/2021.

Trong [6], các tác giả bằng cách sử dụng các nguồn có khuyết điểm đã chỉ ra sự tương quan giữa kiểm tra *entropy xấp xỉ*, *nối tiếp xếp chồng* (Overlapping serial) và kiểm tra *thống kê phổ quát* (Universal).

Trong [12], Turan cùng cộng sự khi xem xét các dãy mẫu có độ dài ngắn $n = 20$ và $n = 30$ đã chỉ ra rằng kiểm tra *tần số*, kiểm tra *so khớp mẫu chồng lấp* (đối với mẫu đầu vào 111), kiểm tra *độ dài run dài nhất của* 1, *chiều cao bước ngẫu nhiên* và *độ phức tạp bậc lớn nhất* là có tính tương quan (thông qua việc quan sát mọi dãy có thể). Bên cạnh đó, Turan cùng cộng sự cũng đề xuất khái niệm về *độ nhạy của các kiểm tra đối với các phép biến đổi đơn giản*.

Doğnaksoy và cộng sự [3], [4] đã áp dụng cách tiếp cận mới và mở rộng các ý tưởng trong [12] tới các dãy có độ dài lớn hơn để đánh giá tính độc lập và độ nhạy của một số kiểm tra thống kê được chọn từ Bộ kiểm tra NIST.

Georgescu và cộng sự [5] đã tiến thêm một bước nữa, tập trung vào các câu hỏi mở liên quan đến tính độc lập của các kiểm tra thống kê của NIST SP800-22, cụ thể nhóm của Georgescu và cộng sự đã nhận xét về độ phức tạp của bộ kiểm tra NIST và về các phương pháp kiểm tra tính độc lập. Họ phát hiện ra rằng, kích thước mẫu có ảnh hưởng đáng kể đến tính tương quan của các thống kê kiểm tra khác nhau (tương tự các quan sát như của nhóm Turan và cộng sự). Họ cũng nhận xét rằng, các kiểm tra tốt hơn có thể tồn tại và việc xác định các kiểm tra đó là rất cần thiết trong nghiên cứu thuộc lĩnh vực này.

Gần đây nhất, vào năm 2019, J.A. Karell-Albo cùng cộng sự [7] đã đề xuất một phương pháp nhằm phát hiện sự phụ thuộc thống kê giữa các phép kiểm tra trong NIST bằng cách sử dụng *thông tin tương hỗ*. Ưu điểm chính của việc sử dụng thông tin tương hỗ là khả năng phát hiện các mối tương quan phi tuyến, điều mà phương pháp hệ số tương quan tuyến tính Pearson đã không thể phát hiện được.

Đóng góp của bài báo: Trong bài báo này, chúng tôi trình bày một cách hệ thống các phương pháp đánh giá sự tương quan giữa các phép kiểm tra thống kê nói chung và đưa ra các kết quả đánh giá thực nghiệm cho bộ kiểm tra NIST SP800-22. Cụ thể, chúng tôi đã đánh giá tính độc lập của một số phép kiểm tra thông qua một số phương pháp bao gồm: *hệ số tương quan Pearson*, *tỷ số Fail-Fail (FF)*, *thông tin tương hỗ*, *độ bao phủ và độ hiệu quả bao phủ*). Kết quả thực nghiệm của chúng tôi cho thấy phát hiện thêm mối tương quan mới giữa các phép kiểm tra thống kê của NIST SP 800-22 khi so sánh với các kết quả nghiên cứu trước đây.

Bố cục của bài báo: Phần II trình bày các kiến thức cơ bản về hệ số tương quan Pearson, thông tin tương hỗ, một số phép biến đổi cơ bản và các phép kiểm tra thống kê trong NIST SP800-22. Nội dung chính tập trung ở Phần III và Phần IV, trong đó chúng tôi trình bày cơ cở lý thuyết cho việc đánh giá mối tương quan giữa các phép kiểm tra thống kê ở Phần III và phân tích các kết quả thực nghiệm để đưa ra một số phân tích, nhận xét ở Phần IV. Cuối cùng, kết luận được trình bày ở Phần V.

II. Kiến thức chuẩn bị

A. Hệ số tương quan Pearson

Giá trị của hệ số tương quan Pearson là một trị số thống kê dùng để đo lường độ mạnh và chiều hướng liên hệ tương quan *tuyến tính* giữa hai biến ngẫu nhiên.

Định nghĩa 1. ***(Hệ số tương quan Pearson)*** *Gọi X và Y là hai biến ngẫu nhiên. Khi đó, hệ số tương quan Pearson giữa hai biến ngẫu nhiên X và Y, ký hiệu là r_{xy}, được cho bởi công thức:*

$$r_{xy} = Corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}, \qquad (1)$$

trong đó $cov(X,Y)$ là hiệp phương sai giữa X và Y, và σ_X, σ_Y tương ứng là độ lệch chuẩn của X và Y.

Hệ số tương quan Pearson, khi được áp dụng cho một mẫu, có thể được biểu diễn như sau:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

trong đó, $\{(x_1, y_2), \ldots, (x_n, y_n)\}$ là n cặp dữ liệu cho trước và n được gọi là cỡ mẫu.

Ý nghĩa giá trị hệ số Pearson: Hệ số tương quan Pearson r_{xy} có giá trị thuộc khoảng $[-1, 1]$, trong đó:

- $r_{xy} = 0$: cho thấy hai biến X và Y không có mối liên hệ tương quan tuyến tính.
- $r_{xy} \in \{1, -1\}$: cho thấy hai biến X và Y có mối tương quan tuyến tính tuyệt đối.
- $r_{xy} < 0$: biểu thị *mối tương quan nghịch*, nghĩa là giá trị các biến X, Y *biến thiên ngược chiều*.
- $r_{xy} > 0$: thể hiện *mối tương quan thuận*, nghĩa là giá trị các biến biến X, Y *biến thiên cùng chiều*.

BẢNG I: MỐI TƯƠNG QUAN ĐỐI VỚI HỆ SỐ PEARSON

$\mathbf{r_{xy}}$	**Mối tương quan**
$(-1, -0.5) \cup (0.5, 1)$	Mạnh
$(-0.49, -0.3) \cup (0.3, 0.49)$	Trung bình
$(-0.29, 0) \cup (0, 0.29)$	Yếu

B. Thông tin tương hỗ

Thông tin tương hỗ (*Mutual Information* - MI) là một đại lượng đo mối quan hệ giữa hai biến ngẫu nhiên được lấy mẫu đồng thời. Cụ thể, thông tin tương hỗ đo mức độ trung bình của lượng thông tin được truyền trong một biến ngẫu nhiên thông qua giá trị của một biến ngẫu nhiên khác. Nói cách khác, độ đo này xác định, về mặt trung bình, sự thay đổi của phân bố của biến ngẫu nhiên X khi biết giá trị biến ngẫu nhiên Y.

Định nghĩa 2 (**Thông tin tương hỗ**, Định nghĩa 2.3, Trg.19, [2])**.** *Thông tin tương hỗ (MI) của hai biến ngẫu nhiên X và Y là một độ đo phản ánh mối quan hệ giữa hai biến này. Khi X và Y là các biến ngẫu nhiên rời rạc, thông tin tương hỗ MI được định nghĩa như sau:*

$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (2)$$

trong đó, $p(x)$ và $p(y)$ tương ứng là hàm phân phối biên duyên của X và Y; và $p(x,y)$ là hàm phân phối xác suất đồng thời của X và Y.

Thông tin tương hỗ được biểu diễn thông qua *entropy* như sau:

$$I(X,Y) = H(X) + H(Y) - H(X,Y)$$

trong đó, $H(X)$, $H(Y)$, và $H(X,Y)$ tương ứng là entropy của X, Y, và (X,Y) được cho bởi:

$$H(X) = -\sum_{x \in X} p(x) \log p(x);$$
$$H(Y) = -\sum_{y \in Y} p(y) \log p(y);$$

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y)$$

Thông tin tương hỗ có các tính chất sau:

1) $I(X,Y) = I(Y,X)$
2) $I(X,X) = H(X)$

C. Bộ kiểm tra thống kê của NIST SP 800-22

Bộ kiểm tra thống kê của NIST SP 800-22 [1] bao gồm tổng cộng 15 phép kiểm tra để đánh giá tính ngẫu nhiên của một dãy hoặc một bộ sinh số ngẫu nhiên cho trước. Bảng II mô tả ký hiệu và viết tắt của 17 phép kiểm tra thống kê trong NIST SP 800-22 (bởi vì có hai phiên bản của *phép kiểm tra tổng tích lũy* và hai phiên bản của phép *phép kiểm tra Serial*).

III. ĐÁNH GIÁ VỀ MỐI TƯƠNG QUAN THỐNG KÊ GIỮA CÁC PHÉP KIỂM TRA

Trong phần này, chúng tôi trình bày cơ sở lý thuyết cho một số phương pháp đánh giá mỗi tương quan giữa các kiểm tra thống kê đã được sử dụng trong các nghiên cứu cho đến nay. Trước tiên, ta nhắc lại khái niệm tính độc lập giữa 2 phép kiểm tra thống kê. Hai phép kiểm tra T_1 và T_2 được gọi là *độc lập* nếu phân bố của các giá trị thống kê (hoặc các trị số p-value) tương ứng của các phép kiểm tra là độc lập. Cụ thể, trong bài báo này, chúng tôi xem xét phân bố của các trị số p-value tương ứng với các phép kiểm tra để đánh giá tính độc lập giữa các phép kiểm tra. Trong phần còn lại của bài báo, chúng tôi ký hiệu $T_i(s_j)$ là trị số p-value của phép kiểm tra T_i đối với dãy s_j.

BẢNG II: KÝ HIỆU VÀ VIẾT TẮT CỦA CÁC PHÉP KIỂM TRA THỐNG KÊ TRONG NIST SP 800-22

Phép kiểm tra	Ký hiệu	Phép kiểm tra	Ký hiệu
Tần số (*Freq*)	T1	Entropy xấp xỉ (*App Ent*)	T10
Tần số trong khối (*Bl Freq*)	T2	Tổng tích lũy về phía trước (*Cusum (f)*)	T11
Run	T3	Tổng tích lũy về phía sau (*Cusum (b)*)	T12
Run dài nhất của 1 (*Long Run*)	T4	So khớp mẫu không chồng lấp (*Non Overl*)	T13
Hạng ma trận (*Rank*)	T5	So khớp mẫu chồng lấp (*Overl*)	T14
Biến đổi Fourier rời rạc (*DFT*)	T6	Độ phức tạp tuyến tính (*LC*)	T15
Phổ quát của Maurer (*Uni Sta*)	T7	Chu trình ngẫu nhiên (*REx*)	T16
Serial 1 (*Ser1*)	T8	Biến thể chu trình ngẫu nhiên (*VREx*)	T17
Serial 2 (*Ser2*)	T9		

Phương pháp đánh giá hệ số Pearson

- Xét các phép kiểm tra T_1, T_2 đối với một tập các dãy $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$ và đưa ra $2n$ các trị số p-value tương ứng $(T_1(\mathbf{s}_1), T_1(\mathbf{s}_2), \ldots, T_1(\mathbf{s}_n))$, $(T_2(\mathbf{s}_1), T_2(\mathbf{s}_2), \ldots, T_2(\mathbf{s}_n))$.
- Tính hệ số tương quan Pearson dựa trên tập dữ liệu gồm n cặp trị số p-value $(T_1(\mathbf{s}_i), T_2(\mathbf{s}_i))$ $(i = 1, \ldots, n)$ ở trên. Dựa vào giá trị hệ số tương quan Pearson được tính để đưa ra đánh giá và quyết định xem T_1 và T_2 có mối tương quan hay không.

A. Phương pháp hệ số tương quan Pearson

Trong mục này, chúng tôi phân tích sự tương quan của một số phép kiểm tra thống kê của NIST SP 800-22, bằng cách xem xét phân bố p-value của các phép kiểm tra đối với cùng một tập hợp các dãy mẫu và tính giá trị hệ số tương quan Pearson giữa các phép kiểm tra thống kê thông qua các trị số p-value. Phương pháp hệ số tương quan Pearson được đề xuất và trình bày trong tài liệu [3], [4].

B. Phương pháp sử dụng Tỷ số Fail-Fail

Hệ số tương quan Pearson chỉ có thể phát hiện các tương quan tuyến tính. Do vậy, chúng ta cần xác định thêm các độ đo khác để xác định các mối tương quan khác (phi tuyến) giữa các phép kiểm tra. Dưới đây, chúng tôi trình bày về *Tỷ số Fail-Fail (FF)* cho phép phát hiện mối tương quan giữa các phép kiểm tra.

Một phép kiểm tra tính ngẫu nhiên sẽ xét xem liệu dãy cần kiểm tra có vượt qua phép kiểm tra hay không, tức là xem dãy có *đạt* (Pass) hay *không đạt* (Fail) tính ngẫu nhiên. Căn cứ vào tỷ lệ giữa số các dãy không đạt cả 2 tiêu chuẩn so với số dãy không đạt của từng tiêu chuẩn có thể rút ra môt kết luận nào đó về mối tương quan giữa hai phép kiểm tra.

Ma trận P-F: Xét một tập các dãy ngẫu nhiên, tương ứng là 2 tập các giá trị p-value tương ứng với 2 phép kiểm tra. Khi đó, ta xây dựng một ma trận gọi là *ma trận P-F* với các phần tử vị trí (i, j) là 0 nếu trị số p-value $T_j(\mathbf{s}_i) < \alpha = 0.01$ (tức là dãy s_i *không đạt* đối với kiểm tra T_j). Tương tự, phần tử vị trí (i, j) là 1 nếu $T_j(\mathbf{s}_i) > \alpha = 0.01$, tức là dãy s_i *đạt* đối với kiểm tra T_j.

C. Phương pháp sử dụng độ hiệu quả bao phủ

Sự phụ thuộc giữa hai phép kiểm tra còn được xác định theo khái niệm độ bao phủ. Để định nghĩa độ bao phủ, trước tiên chú ý rằng, với mức ý nghĩa $\alpha = 0.01$ được cố định, ta nói rằng dãy $\mathbf{s}$ "*không đạt*" đối với phép kiểm tra T_i nếu trị số p-value $T_i(\mathbf{s})$ nhỏ hơn 0.01. Ta có khái niệm về độ bao phủ của một bộ test như sau.

Định nghĩa 3 (Độ bao phủ, [11]**).** *Độ bao phủ của bộ kiểm tra được định nghĩa là tỷ lệ giữa số lượng tất cả các dãy không đạt đối với ít nhất một trong các phép kiểm tra trong bộ kiểm tra và với toàn bộ số dãy mẫu.*

Một tập dữ liệu có thể là ngẫu nhiên theo một số phép kiểm tra, nhưng có thể không ngẫu nhiên theo một phép kiểm tra khác. Do đó, *độ bao phủ* của các kiểm tra được sử dụng trong bộ kiểm

Phương pháp đánh giá tỷ số Fail-Fail

Từ tập các trị số p-value, ta xây dựng một *ma trận P-F* và tiến hành đánh giá như sau:

- Ta cố định hai phép kiểm tra T_i, T_j và gọi $\mathcal{S}$ là tập hợp các dãy mẫu. Gọi $F_{T_i} = \{\mathrm{s} \in \mathcal{S} \mid T_i(\mathrm{s}) < 0.01\}$ và $F_{T_j} = \{\mathrm{s} \in \mathcal{S} \mid T_j(\mathrm{s}) < 0.01\}$ tương ứng là tập hợp gồm tất cả các dãy không đạt đối với phép kiểm tra T_i và T_j. Khi đó, tương quan giữa T_i và T_j là tỷ lệ của các dãy trong tập hợp F_{T_i} cũng không đạt đối với phép kiểm tra T_j và cho bởi (tương tự đối với tương quan giữa T_j và T_i)

$$Cor(T_i, T_j) = \frac{|F_{T_i} \cap F_{T_j}|}{|F_{T_i}|}, \qquad Cor(T_j, T_i) = \frac{|F_{T_i} \cap F_{T_j}|}{|F_{T_j}|} \tag{3}$$

- Theo (3), ta sẽ xây dựng được một bảng *Tỷ số FF* để tìm sự phụ thuộc lẫn nhau giữa các phép kiểm tra (Bảng Tỷ số FF là không đối xứng bởi phép toán trong (3) không có tính chất giao hoán).
- Nếu giá trị Tỷ số FF $Corr(T_i, T_j) > 0.05$ và $Corr(T_j, T_i) > 0.05$ thì ta có thể xem xét 2 phép kiểm tra T_i và T_j có sự phụ thuộc vào nhau.

tra phải đủ cao. Tuy nhiên, việc bao gồm cả các kiểm tra phụ thuộc có thể dẫn đến kết luận sai về tập dữ liệu.

Với mức ý nghĩa α cho trước, Định lý 1 dưới đây cho ta tính độ bao phủ kỳ vọng theo lý thuyết của tập gồm k phép kiểm tra thống kê.

Định lý 1 (Định lý 1, [11]). *Gọi $\mathcal{T}(k)$ là bộ test gồm k phép kiểm tra thống kê độc lập. Gọi α là mức ý nghĩa của k kiểm tra thống kê độc lập. Khi đó, độ bao phủ kỳ vọng (theo lý thuyết) của tập k phép kiểm tra này là:*

$$Cov_{lt}(\mathcal{T}(k)) = \sum_{i=1}^{k} (-1)^{i+1} \binom{k}{i} \alpha^i = 1 - (1-\alpha)^k$$

Chứng minh. Gọi T_i $(i = 1, \ldots, k)$ là k phép kiểm tra thống kê độc lập. Gọi A_i là biến cố biểu thị một dãy không đạt khi được thực hiện qua phép kiểm tra T_i đối với toàn bộ dãy mẫu. Khi đó, ta có độ bao phủ của tập gồm k phép kiểm tra T_i $(i = 1, \ldots, k)$ là $\Pr\left[\bigcup_{i=1}^{k} A_k\right]$. Theo nguyên lý bao hàm-loại trừ (*inclusion-exclusion principle*), ta có:

$$\begin{aligned}\Pr\Big[\bigcup_{i=1}^{k} A_k\Big] &= \sum_{i=1}^{k} \Pr[A_i] - \sum_{1 \le i < j \le k} \Pr[A_i \cap A_j] + \\ &\quad \cdots + (-1)^{k-1} \Pr[A_1 \cap A_2 \cap \ldots \cap A_k] \\ &= \sum_{i=1}^{k} (-1)^{i+1} \Big(\sum_{1 \le j_1 < j_2 < \ldots < j_i \le k} \Pr[A_{j_1} \cap \ldots \cap A_{j_i}] \Big).\end{aligned}$$

Ta thấy, có $\binom{k}{i}$ số hạng bằng nhau trong phép lấy tổng thứ i^{th} của công thức ở trên. Do đó:

$$\begin{aligned}\Pr\Big[\bigcup_{i=1}^{k} A_k\Big] &= \binom{k}{1} \Pr[A_1] - \binom{k}{2} \Pr[A_1 \cap A_2] + \\ &\quad \cdots + (-1)^{k-1} \binom{k}{k} \Pr[A_1 \cap A_2 \cap \cdots \cap A_k] \\ &= \sum_{i=1}^{k} (-1)^{i+1} \binom{k}{i} \Pr[A_1 \cap \ldots \cap A_i] \qquad (4)\end{aligned}$$

Vì k phép kiểm tra thống kê là độc lập, nên ta thu được:

$$\Pr[A_1 \cap A_2 \cap \ldots \cap A_i] = \prod_{j=1}^{i} \Pr[A_j] = \alpha^i \tag{5}$$

Áp dụng đẳng thức (5) vào công thức (4), ta có:

$$\Pr\Big[\bigcup_{i=1}^{k} A_k\Big] = \sum_{i=1}^{k} (-1)^{i+1} \binom{k}{i} \alpha^i = 1 - (1-\alpha)^k$$

□

Áp dụng Định lý 1, ta tính giá trị độ bao phủ kỳ vọng theo lý thuyết của k kiểm tra độc lập với $k \in \{1, 2, \ldots, 17\}$, và được cho ở Bảng III.

Tiếp theo, chúng tôi trình bày về khái niệm *độ hiệu quả bao phủ* của một bộ kiểm tra gồm k phép kiểm tra thống kê.

Định nghĩa 4 (Độ hiệu quả bao phủ, [11]). *Độ hiệu quả bao phủ của một bộ kiểm tra gồm k phép kiểm tra thống kê tính ngẫu nhiên là tỷ số giữa độ bao phủ của bộ kiểm tra với độ bao phủ kỳ vọng theo lý thuyết.*

BẢNG III: TỶ SỐ ĐỘ BAO PHỦ KỲ VỌNG THEO LÝ THUYẾT ĐỐI VỚI $\alpha = 0.01$

Số kiểm tra	Độ bao phủ	Số kiểm tra	Độ bao phủ
1	0.010000	10	0.095618
2	0.019900	11	0.104662
3	0.029701	12	0.113615
4	0.039404	13	0.122479
5	0.049010	14	0.131254
6	0.058520	15	0.139942
7	0.067935	16	0.148542
8	0.077255	17	0.157057
9	0.086483		

Đối với một bộ kiểm tra, chúng ta mong muốn bộ kiểm tra sẽ chứa các phép kiểm tra độc lập. Nếu một bộ kiểm tra chứa các phép kiểm tra có sự tương quan với nhau, thì độ hiệu quả bao phủ của bộ kiểm tra này sẽ thấp hơn độ bao phủ kỳ vọng theo lý thuyết. Thông thường, độ hiệu quả bao phủ nhỏ hơn 1 và khi mà độ hiệu quả bao phủ có giá trị càng nhỏ thì cho thấy sự phụ thuộc càng mạnh giữa các phép kiểm tra.

D. Phương pháp sử dụng thông tin tương hỗ

Để phát hiện sự tương quan giữa các phép kiểm tra thống kê, việc áp dụng cho hai phép kiểm tra được mô hình hóa như một *kênh nhị phân không đối xứng*, sao cho nếu hai phép kiểm tra là tương quan với nhau, thì thông tin tương hỗ của kênh sẽ khác 0 và là một hàm tăng theo bậc tương quan giữa các phép kiểm tra. Phương pháp này sử dụng *thông tin tương hỗ* để phát hiện sự phụ thuộc giữa các phép kiểm tra thống kê. Phương pháp này được đề xuất bởi J.A. Karell-Albo cùng cộng sự [7] và được mô tả theo các bước sau:

- Gọi $\mathcal{S}$ là tập gồm n dãy mẫu $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Xét k phép kiểm tra thống kê T_j $(j = 1, \ldots, k)$ và tính các trị số p-value cho từng dãy đối với mỗi phép kiểm tra T_i. Ký hiệu p_j^i là trị số p-value của dãy $\mathbf{s}_i$ đối với phép kiểm tra T_j $(i = 1, \ldots, n$ và $j = 1, \ldots, k)$.
- Tính các *thông tin tương hỗ* $I(T_i, T_j)$ giữa các phép kiểm tra T_i, T_j dựa theo các dãy trị số p-value: $\{p_j^1, p_j^2, \ldots, p_j^n\}_{j=1,\ldots,k}$.
- Ước tính thông tin tương hỗ $I(T_i, T_j)$ giữa tất cả các cặp (T_i, T_j) của dãy trị số p-value, để phát hiện sự hiện diện của mối tương quan. Việc ước tính thông tin tương hỗ bằng cách sử dụng công thức tính thông tin tương hỗ dựa trên entropy H và một *hàm ước lượng* của H với *sai số toàn phương trung bình* (Mean Squared Error) thấp trên n dãy mẫu.
- Xác định các mối tương quan có ý nghĩa để kết luận mối tương quan giữa các phép phép kiểm tra. Do tính chất đối xứng của thông tin tương hỗ, các giá trị thông tin tương hỗ $I(T_i, T_j)$ được biểu diễn dưới dạng ma trận tam giác trên M như sau:

$$M = \begin{pmatrix} I(T_1,T_1) & I(T_1,T_2) & \ldots & I(T_1,T_k) \\ 0 & I(T_2,T_2) & \ldots & T(T_2,T_k) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & I(T_k,T_k) \end{pmatrix}$$

Vì $I(X, X) = H(X)$, nên ta có thể thay thế các phần tử thuộc đường chéo chính và ma trận M được biểu diễn dưới dạng:

$$M = \begin{pmatrix} H(T_1) & I(T_1,T_2) & \ldots & I(T_1,T_k) \\ 0 & H(T_2) & \ldots & T(T_2,T_k) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & H(T_k) \end{pmatrix}_{k\times k}$$

trong đó, $H(T_i)$ là entropy của biến T_i. Để giải thích tốt hơn, chúng ta thực hiện *chuẩn hóa* các giá trị $I(T_i, T_j)$ bởi $I'(T_i, T_j)$. Có ba biến thể thường dùng để chuẩn hóa thông tin tương hỗ giữa hai biến T_i và T_j là: *chia cho entropy cực đại*, $\max\{H(T_i), H(T_j)\}$, *chia cho entropy nhỏ nhất*, $\min\{H(T_i), H(T_j)\}$ và *chia cho giá trị entropy trung bình*, $(H(T_i) + H(Tj))/2$. Nếu T_i và T_j có cùng phân bố, thì $H(T_i) = H(T_j)$, khi đó ba biến thể ở trên trùng nhau và có giá trị bằng với hệ số:

$$I'(T_i, T_j) = \frac{I(T_i, T_j)}{H(T_i)}$$

Phương pháp đánh giá độ bao phủ và độ hiệu quả bao phủ

- Với một tập hợp các dãy mẫu và các phép kiểm tra thống kê cho trước, ta tính các trị số p-value của các dãy đối với các phép kiểm tra và qua đó thiết lập một ma trận P-F.
- Gọi $\mathcal{T}$ là một bộ gồm k phép kiểm tra T_i $(i = 1, \ldots, k)$ và $F_{T_i} = \{\mathbf{s} \in \mathcal{S} \mid T_i(\mathbf{s}) < 0.01\}$ là tập các dãy không đạt đối với kiểm tra T_i, trong đó $\mathcal{S}$ là không gian mẫu. Khi đó, **độ bao phủ** của $\mathcal{T}$:

$$Cov(\mathcal{T}) = \frac{|\bigcup_i F_{T_i}|}{|\mathcal{S}|} \tag{4}$$

  Với độ bao phủ tính được, ta sẽ so sánh với giá trị độ bao phủ kỳ vọng theo lý thuyết. Nếu độ bao phủ của bộ kiểm tra càng gần với độ bao phủ kỳ vọng theo lý thuyết, thì tính độc lập giữa các phép kiểm tra càng tốt. Nếu giá trị độ bao phủ nhỏ hơn giá trị độ bao phủ kỳ vọng theo lý thuyết, thì đó là dấu hiệu cho biết các phép kiểm tra có mối tương quan hoặc yếu trong bộ kiểm tra.
- Tính **độ hiệu quả bao phủ**

$$Eff - Cov(\mathcal{T}) = \frac{Cov(\mathcal{T})}{Cov_{theo}(\mathcal{T}(k))} \tag{5}$$

Ta thấy rằng, bởi vì tập các dãy mẫu được sinh từ một bộ sinh số giả ngẫu nhiên, T_i và T_j có cùng phân bố và do đó các trị số p-value phân phối đều. Khi đó, ta có:

$$M = \begin{pmatrix} 1 & I'(T_1, T_2) & \ldots & I'(T_1, T_k) \\ 0 & 1 & \ldots & T'(T_2, T_k) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}_{k \times k}$$

trong đó, $I'(T_i, T_j)$ là giá trị *thông tin tương hỗ được chuẩn hóa* (NMI) giữa các biến T_i và T_j.

- Quyết định xem $I'(T_i, T_j)$ có lớn hơn 0 đáng kể hay không và từ đó kết luận liệu có sự phụ thuộc nào giữa cả hai biến hay không.

Ta hình thức hóa hai giả thuyết thống kê sau:

$$H_0 : I'(T_i, T_j) = 0; \quad H_1 : I'(T_i, T_j) > 0$$

trong đó, H_0 là *giả thuyết không* mà phát biểu là "*hai kiểm tra có tính độc lập*" và H_1 là giả thuyết nghịch với khẳng định "*có sự phụ thuộc nào đó giữa hai phép kiểm tra* T_i và T_j". Từ tập các trị số p-value mẫu $(p_i^1, p_j^1), \ldots, (p_i^n, p_j^n)$, ta cần quyết định có chăng bác bỏ giả thuyết không H_0 về tính độc lập của hai kiểm tra T_i và T_j hay không. Tức là, nếu $I'(T_i, T_j)$ là lớn hơn đáng kể 0, thì giả thuyết H_0 bị bác bỏ. Đối với điều này, cần phải tính trị số p-value liên quan đến phép kiểm tra và nếu trị số p-value này nhỏ hơn mức ý nghĩa đã chọn α thì sẽ có đủ bằng chứng bác bỏ H_0.

Vì không biết phân bố của thông tin tương hỗ dưới giả thuyết H_0, nên sẽ thực hiện một kiểm tra hoán vị như sau:

1) Tạo q mẫu hoán vị $\{\pi_u(T_j)\}_{u=1,\ldots,q}$ của phép kiểm tra T_j. Xây dựng các mẫu được hoán vị $(T_i, \pi_u(T_j))$ giữa 2 phép kiểm tra T_i, T_j, trong đó hoán vị π_u là hoán vị thứ u của các trị số p-value của phép kiểm tra T_j sao cho thỏa mãn 3 điều kiện sau:
   - $\pi_u \in S_n$ với mọi $u = 1, \ldots, q$, trong đó S_n ký hiệu là tập hợp tất cả các hoán vị của các trị số p-value của T_j.
   - $\pi_u \neq \pi_v$ với mọi $u \neq v$.
   - π_0 là hoán vị đồng nhất của S_n, tức là $\pi_0(T_j) = T_j$.
2) Ước lượng các thông tin tương hỗ được chuẩn hóa $Z_0 = I'(T_i, T_j)$ và $Z_u = I'(T_i, \pi_u(T_j))$ với $u = 1, \ldots, q$.
3) Tính trị số p-value (đối với kiểm tra hoán vị) dựa theo thông tin tương hỗ ở bước 2.

$$p - value = \frac{\sum_{u=1}^{q} \mathbb{1}_{\geq Z_0}(Z_u)}{q}$$

trong đó, $\mathbb{1}_{\geq Z_0}(Z_u)$ là hàm chỉ thị cho bởi:

$$\mathbb{1}_{\geq Z_0}(x) = \begin{cases} 1 & \text{nếu } x \geq Z_0 \\ 0 & \text{nếu } x < Z_0 \end{cases}.$$

4) Nếu $p - value \geq \alpha$, thì giả thuyết H_0 được chấp nhận, tức là hai phép kiểm tra T_i, T_j là có tính độc lập. Nếu $p - value < \alpha$, thì giả thuyết H_0 bị bác bỏ, nghĩa là T_i và T_j có sự phụ thuộc nào đó.

Lựa chọn số lượng dãy mẫu và hàm ước lượng thông tin tương hỗ

Vì chúng ta không thể tính chính xác giá trị thông tin tương hỗ nên sẽ cần sử dụng một hàm ước lượng thông tin tương hỗ. Hàm ước lượng phải được lựa chọn một cách cẩn thận vì nếu *hàm ước lượng thông tin tương hỗ* không được chọn tốt thì có thể ảnh hưởng tới độ chính xác của thông tin tương hỗ.

Thông tin tương hỗ được ước lượng dựa vào công thức biểu diễn theo dạng entropy. Ước lượng được thực hiện đối với hai trường hợp: *biến phụ thuộc* và *biến độc lập*. Trong số các công cụ ước lượng entropy, chúng ta chọn trong gói `infotheo` trong *ngôn ngữ lập trình R* (phổ biến trong phân tích, tính toán thống kê). Trong gói `infotheo` bao gồm 4 tùy chọn phương pháp ước lượng: ước lượng (*emp*), ước lượng Miller–Madow (*mm*), ước lượng shrinkage James-Stein (*shrink*) và ước lượng Schurmann-Grassberger (*sg*). Bằng thực nghiệm, J.A.Karell-Albo cùng cộng sự [7] đã chỉ ra rằng sự khác biệt lớn nhất giữa các giá trị ước lượng sẽ giảm đáng kể khi số dữ liệu quan sát tăng trong trường hợp các biến độc lập. Cụ thể hơn, J.A.Karell-Albo cùng cộng sự [7] thấy rằng "*nếu ít nhất* $10,000$ *dãy số giả ngẫu nhiên được phân tích, thì sự khác biệt lớn nhất giữa các bộ ước lượng là rất nhỏ*" và lựa chọn ước lượng (shrink) để tính thông tin tương hỗ.

IV. Kết quả thực nghiệm

A. Thiết lập thí nghiệm

Tập các dữ liệu mẫu được sinh từ *Bộ sinh số giả ngẫu nhiên* HMAC-DRBG với đầu vào là một mầm true random. Tập dữ liệu mẫu gồm $n = 10,000$ dãy có độ dài $L = 1,000,000$ (bit). Đối với mỗi phép kiểm tra T_i $(i = 1, \ldots, 17)$ (trong 17 phép kiểm tra của NIST), chúng tôi tính $10,000$ trị số p-value $(p_i^1, \ldots, p_i^n)$ tương ứng của n dãy mẫu đối với phép kiểm tra T_i và đồng thời tạo $q = 10,000$ hoán vị phân biệt đối với mỗi tập dữ liệu $(p_i^1, \ldots, p_i^n)$.

Bảng IV trình bày các tham số chúng tôi sử dụng trong mỗi phép kiểm tra thống kê theo khuyến cáo của NIST.

B. Phân tích, đánh giá kết quả thực nghiệm

Ta lưu ý rằng khi phép kiểm tra, giả sử là T_1, có mối tương quan với một phép kiểm tra khác, giả sử là T_2 thì T_2 cũng có mối tương quan với T_1. Khi đó, trong các Bảng so sánh kết quả tương quan ở suốt mục này, chúng ta quy ước rằng nếu T_1 được nhắc là có tương quan với T_2 ở dòng trước đó thì khi xét đến T_2 chúng ta bỏ qua việc nhắc lại T_1.

1) Đối với phương pháp hệ số Pearson: Bảng XI trình bày các giá trị hệ số tương quan Pearson giữa các phép kiểm tra thống kê của NIST SP800-22. Về mặt lý thuyết, để các phép kiểm tra không tương quan, thì tất cả các giá trị hệ số tương quan Pearson phải bằng 0. Tuy nhiên, trong thực tế do việc lựa chọn dữ liệu mẫu khiến có thể tạo ra giá trị hệ số Pearson với một số sai lệch so với 0 ngay cả khi các phép kiểm tra là không tương quan. Trong Bảng XI, *các giá trị hệ số tương quan Pearson có giá trị tuyệt đối lớn hơn* 0.5 *sẽ thể hiện cho sự tương quan giữa hai phép kiểm tra*. Bảng V trình bày so sánh kết quả thu được từ thực nghiệm của chúng tôi và kết quả từ bài báo [4].

2) Đối với phương pháp tỷ số FF: Theo Bảng XIII, chúng tôi nhận thấy rằng:

- Phép kiểm tra *tần số* (Freq) có mối tương quan với các phép kiểm tra *tổng tích lũy* CuSum(f,b), *(biến thể) chu trình ngẫu nhiên* (**REx**, **VREx**).
- Các phép kiểm tra *Entropy xấp xỉ*, *Serial-1* và *Serial-2* có mối tương quan với nhau.
- Phép kiểm tra *Tần số trong khối* (Bl Freq) có mối tương quan với phép kiểm tra *tổng tích lũy về phía trước* CuSum(f).

BẢNG IV: CÁC THAM SỐ ĐƯỢC SỬ DỤNG TRONG CÁC PHÉP KIỂM TRA THỐNG KÊ

Phép kiểm tra	**Tham số**	**Giá trị**
Entropy xấp xỉ	m : độ dài khối đầu tiên	8
Tần số trong khối	M : độ dài khối	12500
Độ phức tạp tuyến tính	M : độ dài khối	500
Run dài nhất	M : độ dài khối	10000
So khớp mẫu không chồng lấp	mẫu	000000001
So khớp mẫu chồng lấp	mẫu	11111111
Chu trình ngẫu nhiên	trạng thái bên trong	$x = 1$
Biến thể chu trình ngẫu nhiên	trạng thái bên trong	$x = -9$
Hạng ma trận	(M, Q) : số hàng và số cột của mỗi ma trận	$M = 32, Q = 32$
Serial	m : độ dài khối	8
Thống kê phổ quát của Maurer	L : độ dài khối, Q : các khối khởi tạo	$L = 7$, $Q = 10$

BẢNG V: SO SÁNH CÁC KẾT QUẢ TƯƠNG QUAN THU ĐƯỢC VỚI CÁC NGHIÊN CỨU ĐÃ BIẾT THEO PHƯƠNG PHÁP HỆ SỐ PEARSON [3], [4]. LƯU Ý RẰNG KIỂM TRA IN ĐẬM THỂ HIỆN CHO PHÁT HIỆN THÊM MỐI TƯƠNG QUAN

Kiểm tra	**Các phép kiểm tra tương quan**	
	[3], [4]	***Nhóm tác giả***
Freq	CuSum(f), CuSum(b)	CuSum(f), CuSum(b)
CuSum(f)	CuSum(b)	CuSum(b)
Serial-1	Serial-2	Serial-2
App Ent	Serial-1, Serial-2	Serial-1, Serial-2
REx		**VREx**

BẢNG VI: SO SÁNH CÁC KẾT QUẢ TƯƠNG QUAN THU ĐƯỢC VỚI CÁC NGHIÊN CỨU ĐÃ BIẾT THEO PHƯƠNG PHÁP TỶ SỐ FF. CHÚ Ý RẰNG CÁC PHÉP KIỂM TRA ĐƯỢC PHÁT HIỆN MỚI SẼ ĐƯỢC IN ĐẬM

Kiểm tra	**Các phép kiểm tra có sự tương quan**	
	[8], [11]	***Nhóm tác giả***
Freq	Bl Freq	**CuSum(f, b)**, **REx**, **VREx**
Bl Freq		**CuSum(f)**, **REx**, **VREx**
App Ent	Serial-1, Serial-2	Serial-1, Serial-2, **REx**, **VREx**
Serial-1	Serial-2	Serial-2, **REx**, **VREx**
CuSum(f)		**CuSum(b)**, **REx**, **VREx**
REx		**VREx**, **Run**, **Rank**, **DFT**, **Uni Sta**, **Serial-2**, **Overl**, **Non Overl**, **LC** , **Long Run**, **CuSum(b)**
VREx		**Run**, **Rank**, **DFT**, **Uni Sta**, **Serial-1**, **Overl**, **Non Overl**, **LC** , **Long Run**, **CuSum(b)**

3) Đối với độ bao phủ và độ hiệu quả bao phủ: Bảng X trình bày độ bao phủ và độ hiệu quả bao phủ khi xét 15 phép kiểm tra thống kê (loại trừ 2 phép kiểm tra **REx** (T16) và **VREx** (T17)) với 17 phép kiểm tra thống kê trong NIST SP800-22. Dựa theo kết quả thu được, ta có thể thấy rằng việc bổ sung 2 phép kiểm tra **REx** và **VREx** làm cho độ hiệu quả bao phủ của bộ kiểm tra "*giảm xuống*". Điều này cho thấy dường như có tương quan giữa hai phép kiểm tra **REx** và **VREx** với các phép kiểm tra từ T1 đến T15. Ngoài ra, với việc thử nghiệm đánh giá cho 17 phép kiểm tra thống kê trên tập dữ liệu mẫu, chúng tôi phát hiện thêm mới mối tương quan giữa một số phép kiểm tra và được trình bày ở Bảng VI (trong ngữ cảnh so sánh với kết quả từ các nghiên cứu đã biết). Đặc biệt, ta thấy mối tương quan giữa 2 phép kiểm tra **REx** (T16) và

BẢNG VII: SO SÁNH SỰ TƯƠNG QUAN GIỮA CÁC PHÉP KIỂM TRA THEO THÔNG TIN TƯƠNG HỖ (ĐỐI VỚI $\alpha = 0.001$), TRONG ĐÓ CÁC PHÉP KIỂM TRA IN ĐẬM THỂ HIỆN PHÁT HIỆN THÊM SO VỚI KẾT QUẢ TỪ [7].

Kiểm tra	**Các phép kiểm tra có sự tương quan**	
	[7]	***Nhóm tác giả***
App Ent	Serial-1	Serial-1, **Serial-2**, **REx**, **VREx**
CuSum (f)	CuSum (b), Freq, REx, VREx	CuSum (b), Freq, **Bl Freq**, REx, VREx
CuSum (b)	Freq, REx, VREx	Freq, **Bl Freq**, REx, VREx
Freq	REx, VREx	**Bl Freq**, REx, VREx
Long Run	Overl	Overl, **REx**, **VREx**
REx	VREx	VREx, **Run**, **Rank**, **DFT**, **Uni Sta**, **Serial-2**, **Overl**, **Non Overl**, **LC**
VREx		**Run**, **Rank**, **DFT**, **Uni Sta**, **Serial-2**, **Overl**, **Non Overl**, **LC**
Serial-1	Serial-2	Serial-2, **REx**, **VREx**

VREx (T17) với toàn bộ 15 phép kiểm tra từ T1-T15. Điều này giải thích tại sao độ hiệu quả bao phủ có sự "*giảm xuống*" rõ rệt khi thêm hai phép kiểm tra này.

4) Đối với phương pháp thông tin tương hỗ: Bảng XII biểu diễn các giá trị NMI giữa 17 phép kiểm tra trong Bộ kiểm tra của NIST. Chúng ta có thể thấy phép kiểm tra nào có giá trị NMI lớn hơn $\alpha = 0.001$ sẽ được thể hiện tô màu tương ứng ở Bảng XII. Kết hợp với thực nghiệm phương pháp kiểm tra hoán vị, ta có thể đưa ra bằng chứng nhằm bác bỏ giả thuyết không H_0. Điều này cho thấy xuất hiện sự phụ thuộc giữa một vài phép kiểm tra. Bảng VII cho thấy một so sánh về mối tương quan được phát hiện giữa các phép kiểm tra trong Bộ kiểm tra của NIST.

Nhận xét 1. *Từ kết quả thực nghiệm, chúng tôi phát hiện thêm về sự tương quan giữa các phép kiểm tra thống kê so với kết quả từ bài báo [7] của J.A. Karell-Albo cùng cộng sự. Các phát hiện thêm này được thể hiện dưới dạng in đậm như trong Bảng VII. Đặc biệt, các giá trị NMI giữa hai phép kiểm tra* **REx** *và* **VREx** *với 15 phép kiểm tra còn lại càng củng cố thêm bằng chứng cho thấy rằng có sự tương quan giữa 2 phép kiểm tra này với 15 phép kiểm tra còn lại.*

C. Thảo luận thêm về sự tương quan giữa các trị số p-value

Quan sát biểu đồ phân tán của dãy trị số *p*-value: Một thể hiện khá thú vị là chúng ta có thể quan sát các trị số p-value của các phép kiểm tra thông qua *biểu đồ phân tán*.

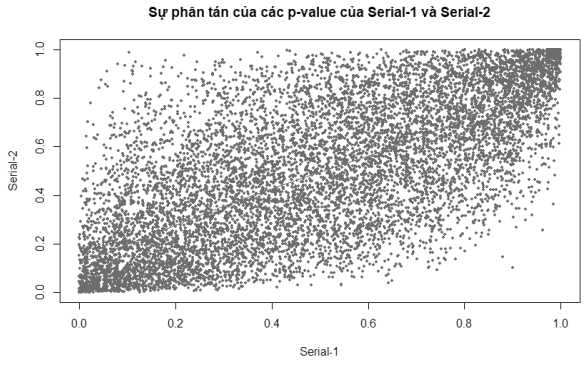
Ví dụ 1. *Xét hai phép kiểm tra: Entropy xấp xỉ và Hạng ma trận mà theo NMI chỉ ra không có mối tương quan. Khi đó, vẽ biểu đồ phân tán các trị số p-value của hai phép kiểm tra này (xem Hình 2), thì ta cũng quan sát thấy không có thể hiện nào cho thấy có sự phụ thuộc giữa 2 phép kiểm tra này.*

Ví dụ 2. *Nếu chúng ta phân tích biểu đồ phân tán của các cặp phép kiểm tra khác, có thể nhận thấy rằng khi giá trị NMI tăng lên, thì dãy trị số p-value phản ánh các biểu hiện liên quan đến sự tương quan phụ thuộc. Điều này được chứng minh bằng các phép kiểm tra Serial-1 và Serial-2 (xem Hình 1a), phép kiểm tra tần số và phép kiểm tra CuSum (b) (xem Hình 1b), kiểm tra CuSum (f) và CuSum (b) (xem Hình 1c), kiểm tra Entropy xấp xỉ và Serial-1 (xem Hình 1d) .*

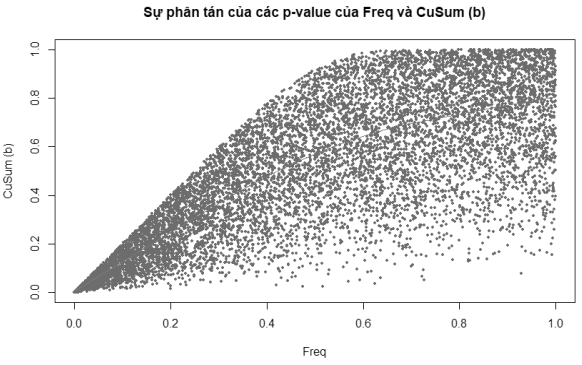
D. So sánh với các kết quả nghiên cứu trước đó

Dựa theo kết quả thực nghiệm, chúng tôi thực hiện so sánh với các kết quả nghiên cứu đã công bố từ các tài liệu tham khảo trước đó.

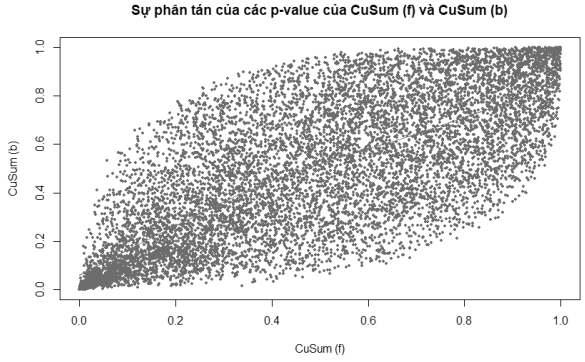
Đối với đánh giá về tính độc lập giữa các phép kiểm tra: Trước tiên, chúng tôi trình bày về môi trường làm việc mà các phương pháp đánh giá thường sử dụng, bao gồm tham số khởi tạo, số lượng dãy mẫu và các tham số được sử dụng bởi mỗi phép kiểm tra, vì khi các tham số này thay đổi sẽ khiến kết quả phép kiểm tra thay đổi đáng kể. Các tham số khởi tạo được sử dụng trong các phương pháp khác nhau được trình bày trong Bảng VIII. Theo khuyến cáo của NIST, nhiều tiêu chuẩn kiểm tra thống kê phải



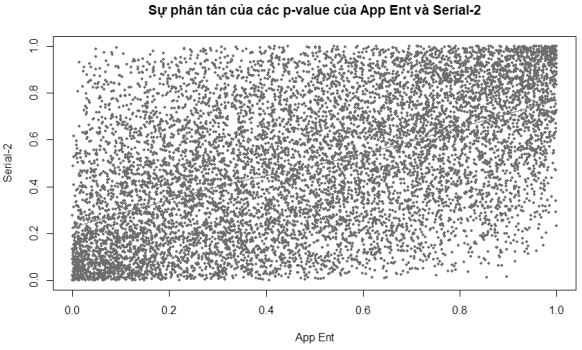
(a) Sự phân tán *p*-value của *Ser-1* và *Ser-2*



(b) Sự phân tán *p*-value của *Freq* và *CuSum (b)*



(c) Phân tán *p*-value của *CuSum(f)* và *CuSum(b)*



(d) Sự phân tán *p*-value của *App Ent* và *Serial-2*

Hình 1. Sự phân tán của các giá trị NMI của một số phép kiểm tra thống kê.

BẢNG VIII: CÁC THAM SỐ VÀ PHƯƠNG PHÁP ĐƯỢC DÙNG TRONG CÁC NGHIÊN CỨU ĐÃ BIẾT

Tài liệu	**Số dãy mẫu n**	**Độ dài dãy L (bit)**	**Phương pháp đánh giá**
[4]	100,000	500	Hệ số tương quan Pearson
[11]	100,000	38,912	Tỷ số Fail-Fail
[3]	200,000 200	1024 1,048,576	Hệ số tương quan Pearson Hệ số tương quan Pearson
[7]	10,000	1,000,000	Thông tin tương hỗ (MI)
Nhóm tác giả	10,000	1,000,000	Hệ số tương quan Pearson Thông tin tương hỗ (MI) Tỷ số Fail-Fail

đáp ứng yêu cầu tham số về độ dài dãy tối thiểu là 10^6 bit, tuy nhiên các kết quả trong [3], [4], [11] chỉ sử dụng các dãy có độ dài ngắn là 500, 38912, 2014 bit, không phù hợp theo khuyến cáo của NIST. Do đó, các công trình [3], [4], [11] đã không sử dụng đúng theo khuyến cáo của NIST, nghĩa là các tác giả trong [3], [4], [11] đã chưa nhận thức đúng về thiết lập tham số nhằm đánh giá kết quả thực nghiệm, tuy nhiên họ vẫn thu được các kết quả mới về tính độc lập của một số tiêu chuẩn kiểm tra thống kê.

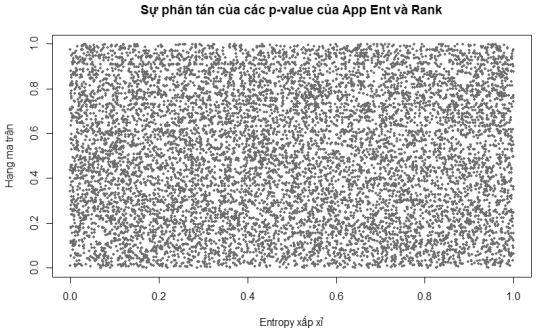
Dựa trên các phương pháp đánh giá đã được đề xuất trong các tài liệu [3], [4], [7], [8], [11], [12], chúng tôi đã thực nghiệm trên tập mới các dữ liệu mẫu, qua đó đưa ra phân tích, đánh giá về tính độc lập, độ nhạy của các phép kiểm tra thống kê của NIST. Kết quả là chúng tôi đã phát hiện thêm sự tương quan mới giữa một số các phép kiểm tra trong Bộ kiểm tra NIST SP 800-22. Kết quả này được trình bày ở Bảng IX.

BẢNG IX: SO SÁNH SỰ TƯƠNG QUAN GIỮA CÁC PHÉP KIỂM TRA THỐNG KÊ CỦA NIST SP800-22 MÀ NHÓM TÁC GIẢ THU ĐƯỢC TỪ THỰC NGHIỆM SO VỚI CÁC NGHIÊN CỨU ĐÃ BIẾT, TRONG ĐÓ CÁC PHÉP KIỂM TRA IN ĐẬM THỂ HIỆN PHÁT HIỆN THÊM SO VỚI CÁC KẾT QUẢ NGHIÊN CỨU ĐÃ CÔNG BỐ

Kiểm tra	**Các phép kiểm tra tương quan**			
	[3], [4]	*[8], [11]*	*[7]*	*Nhóm tác giả*
Freq	CuSum(f,b)	CuSum(f,b)	CuSum(f,b), REx, VREx, Run	CuSum(f,b), REx, VREx, Run, **Bl Freq**
App Ent	Serial(1,2)	Serial(1,2)	Serial-1	Serial(1,2), **REx**, **VREx**
CuSum(f), CuSum(b)	CuSum(b)		CuSum(b), REx, VREx, Run	CuSum(b), REx, VREx, Run, **Bl Freq**
Serial-1	Serial-2	Serial-2	Serial-2	Serial-2, **REx**, **VREx**
REx			VREx	VREx, **Run**, **Rank**, **DFT**, **Uni Sta**, **Serial-2**, **Overl**, **Non Overl**, **LC**
VREx				**Run**, **Rank**, **DFT**, **Uni Sta**, **Serial-2**, **Overl**, **Non Overl**, **LC**
Long Run			Overl	Overl, **REx**, **VREx**

BẢNG X: SO SÁNH ĐỘ BAO PHỦ VÀ ĐỘ HIỆU QUẢ BAO PHỦ GIỮA MỘT SỐ NHÓM PHÉP KIỂM TRA VÀ 17 PHÉP KIỂM TRA THỐNG KÊ CỦA NIST. CHÚ Ý $\mathcal{F}/(Ti, \cdots, Tj)$ KÝ HIỆU SỐ PHÉP KIỂM TRA CÒN LẠI SAU KHI LOẠI ĐI CÁC KIỂM TRA $Ti, \cdots, Tj$ TỪ 17 KIỂM TRA BAN ĐẦU

Các kiểm tra	**Độ bao phủ**	**Độ hiệu quả bao phủ**
T1–T17	0.4569	2.909135
$\mathcal{F}/(T16, T17)$	0.1162	0.830344
$\mathcal{F}/(T1, T17)$	0.450800	3.221335
$\mathcal{F}/(T9, T16, T17)$	0.110800	0.844165
$\mathcal{F}/(T12, T16, T17)$	0.114700	0.873878
$\mathcal{F}/(T9, T12, T16, T17)$	0.109300	0.892398
$\mathcal{F}/(T9, T11, T12, T16, T17)$	0.107900	0.949699
$\mathcal{F}/(T8, T9, T11, T12, T16, T17)$	0.101600	0.970744



Hình 2. Sự phân tán p-value của *App Ent* và *Rank*.

KẾT LUẬN

Trong bài báo này, chúng tôi đã thực hiện phân tích và đánh giá về tính độc lập của các phép kiểm tra thống kê về tính ngẫu nhiên trong NIST SP 800-22. Dựa trên các phương pháp đánh giá thống kê đã biết như: *phương pháp hệ số tương quan Pearson*, *tỷ số FF*, *thông tin tương hỗ*, *độ bao phủ và độ hiệu quả bao phủ* kết hợp với kết quả thực nghiệm để tìm mối tương quan giữa các phép kiểm tra thống kê. Kết quả thu được là chúng tôi đã phát hiện thêm mối tương quan mới giữa một số phép kiểm tra thống kê mà chưa được phát hiện trong các công trình trước đó. Điều này cho thấy có thể giảm số lượng phép kiểm tra thống kê có trong Bộ kiểm tra của NIST SP 800-22 bằng cách bỏ bớt các phép kiểm tra có sự phụ thuộc lẫn nhau. Đặc biệt, chúng tôi đã đánh giá độ hiệu quả bao phủ của bộ kiểm tra trong 2 trường hợp khi sử dụng 2 kiểm tra chu trình ngẫu nhiên và biến thể chu trình ngẫu nhiên và khi không sử dụng 2 kiểm tra này. Theo hiểu biết của chúng tôi, đây là những kết quả đầu tiên về độ hiệu quả bao phủ của bộ kiểm tra NIST SP 800-22. Kết quả cho thấy độ hiệu quả bao

BẢNG XI: HỆ SỐ PEARSON GIỮA CÁC TRỊ SỐ *p*-VALUE CỦA CÁC KIỂM TRA TRONG NIST SP800-22.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17
T1	1	0.084	0.008	0.003	-0.009	-0.006	-0.014	0.064	-0.003	0.043	**0.763**	**0.766**	0.024	0.03	-0.013	0.226	0.229
T2	0.084	1	0.008	-0.016	0.016	0.012	-0.009	0.031	0.011	0.027	0.165	0.168	0.023	0	-0.008	0.047	0.04
T3	0.008	0.008	1	0.013	0.005	0.003	0.017	0.067	0.005	0.054	0.008	0.005	0.012	0.018	0.007	-0.004	0.004
T4	0.003	-0.016	0.013	1	-0.008	0.01	-0.007	0.016	0.006	0.011	0.001	0	0.022	0.092	0.004	-0.008	0.001
T5	-0.009	0.016	0.005	-0.008	1	0	-0.029	0	-0.01	0.006	0	0.002	-0.002	-0.004	-0.003	-0.006	0
T6	-0.006	0.012	0.003	0.01	0	1	0.006	-0.013	-0.012	-0.016	-0.003	-0.01	-0.021	0.001	0.006	-0.002	0.023
T7	-0.014	-0.009	0.017	-0.007	-0.029	0.006	1	0.008	-0.001	0.003	-0.015	-0.01	0.004	-0.002	-0.012	0.001	0.002
T8	0.064	0.031	0.067	0.016	0	-0.013	0.008	1	**0.689**	**0.694**	0.045	0.06	0.017	0.04	-0.019	0.014	0.013
T9	-0.003	0.011	0.005	0.006	-0.01	-0.012	-0.001	**0.689**	1	0.486	-0.014	0.003	-0.001	0.013	-0.014	0.007	0.002
T10	0.043	0.027	0.054	0.011	0.006	-0.016	0.003	**0.694**	0.486	1	0.03	0.039	0.02	0.031	-0.001	0.014	0.009
T11	**0.763**	0.165	0.008	0.001	0	-0.003	-0.015	0.045	-0.014	0.03	1	**0.723**	0.038	0.022	-0.021	0.317	0.318
T12	**0.766**	0.168	0.005	0	0.002	-0.01	-0.01	0.06	0.003	0.039	**0.723**	1	0.032	0.022	-0.025	0.128	0.13
T13	0.024	0.023	0.012	0.022	-0.002	-0.021	0.004	0.017	-0.001	0.02	0.038	0.032	1	0.01	0.011	0.004	0.003
T14	0.03	0	0.018	0.092	-0.004	0.001	-0.002	0.04	0.013	0.031	0.022	0.022	0.01	1	-0.001	0.001	0.005
T15	-0.013	-0.008	0.007	0.004	-0.003	0.006	-0.012	-0.019	-0.014	-0.001	-0.021	-0.025	0.011	-0.001	1	0.001	-0.002
T16	0.226	0.047	-0.004	-0.008	-0.006	-0.002	0.001	0.014	0.007	0.014	0.317	0.128	0.004	0.001	0.001	1	**0.547**
T17	0.229	0.04	0.004	0.001	0	0.023	0.002	0.013	0.002	0.009	0.318	0.13	0.003	0.005	-0.002	**0.547**	1

BẢNG XII: THÔNG TIN TƯƠNG HỖ ĐƯỢC CHUẨN HÓA CỦA 17 PHÉP KIỂM TRA TRONG NIST

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17
T1	1	0.002143	0	0	0	0	0	0	0	0	0.215681	0.213502	0	0.000192	0	0.022487	0.022084
T2	0.002143	1	0	0	0	0	0	0	0	0.000112	0.005138	0.00529	0	0	0	0.005283	0.005275
T3	0	0	1	2.70E-05	6.76E-05	0	0	0.000432	6.03E-05	0.000721	0	0	0	0	0	0.004747	0.004131
T4	0	0	2.70E-05	1	8.50E-05	0	0	0	0	0	5.02E-05	0	0	0.001441	0	0.004704	0.004472
T5	0	0	6.76E-05	8.50E-05	1	0	0	0.000386	0	0	0	0	0.000244	0	0	0.003855	0.004312
T6	0	0	0	0	0	1	0	0	0	0	0	0.000251	0	0	0	0.004826	0.004723
T7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0.004155	0.004857
T8	0	0	0.000432	0	0.000386	0	0	1	0.110025	0.112292	4.30E-05	0	0	0	0	0.00432	0.004553
T9	0	0	6.03E-05	0	0	0	0	0.110025	1	0.045313	0	0	0	0	0	0.00506	0.004168
T10	0	0.000112	0.000721	0	0	0	0	0.112292	0.045313	1	0.000137	0	0	0	0	0.004344	0.004677
T11	0.215681	0.005138	0	5.02E-05	0	0	0	4.30E-05	0	0.000137	1	0.14939	0	0.000134	0	0.041534	0.04128
T12	0.213502	0.00529	0	0	0	0.000251	0	0	0	0	0.14939	1	0	0	0	0.01063	0.009928
T13	0	0	0	0	0.000244	0	0	0	0	0	0	0	1	0	0	0.004258	0.004313
T14	0.000192	0	0	0.001441	0	0	0	0	0	0	0.000134	0	0	1	0	0.004209	0.004254
T15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.004643	0.005129
T16	0.03019	0.007093	0.006374	0.006315	0.005176	0.00648	0.005579	0.005799	0.006794	0.005833	0.055762	0.014271	0.005717	0.00565	0.006233	1	0.2998
T17	0.029649	0.007082	0.005546	0.006005	0.00579	0.006341	0.00652	0.006113	0.005596	0.006279	0.055421	0.013329	0.005791	0.005712	0.006886	0.2998	1

BẢNG XIII: MA TRẬN TỶ SỐ FF GIỮA 17 PHÉP KIỂM TRA THỐNG KÊ CỦA NIST

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17
T1	1.000000	0.040000	0.020000	0.000000	0.000000	0.000000	0.010000	0.040000	0.010000	0.020000	**0.750000**	**0.830000**	0.010000	0.030000	0.010000	**0.690000**	**0.690000**
T2	0.040000	1.000000	0.010000	0.010000	0.000000	0.000000	0.010000	0.000000	0.010000	0.010000	**0.080000**	0.040000	0.000000	0.010000	0.020000	**0.440000**	**0.450000**
T3	0.021277	0.010638	1.000000	0.000000	0.010638	0.000000	0.010638	0.010638	0.000000	0.031915	0.010638	0.031915	0.010638	0.000000	0.000000	**0.340426**	**0.351064**
T4	0.000000	0.008772	0.000000	1.000000	0.008772	0.000000	0.026316	0.008772	0.008772	0.008772	0.008772	0.000000	0.000000	0.035088	0.017544	**0.359649**	**0.394737**
T5	0.000000	0.000000	0.012048	0.012048	1.000000	0.000000	0.024096	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.012048	**0.313253**	**0.301205**
T6	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.021978	0.021978	0.000000	0.000000	0.010989	0.000000	0.000000	0.000000	**0.384615**	**0.395604**
T7	0.008130	0.008130	0.008130	0.024390	0.016260	0.000000	1.000000	0.008130	0.016260	0.000000	0.024390	0.008130	0.000000	0.024390	0.008130	**0.512195**	**0.536585**
T8	0.038835	0.000000	0.009709	0.009709	0.000000	0.019417	0.009709	1.000000	**0.262136**	**0.300971**	0.048544	0.038835	0.009709	0.000000	0.000000	**0.436893**	**0.436893**
T9	0.011236	0.011236	0.000000	0.011236	0.000000	0.022472	0.022472	**0.303371**	1.000000	**0.089888**	0.011236	0.011236	0.000000	0.000000	0.011236	**0.370787**	**0.382022**
T10	0.018349	0.009174	0.027523	0.009174	0.000000	0.000000	0.000000	**0.284404**	**0.073394**	1.000000	0.009174	0.018349	0.000000	0.027523	0.009174	**0.440367**	**0.431193**
T11	**0.750000**	**0.080000**	0.010000	0.010000	0.000000	0.000000	0.030000	0.050000	0.010000	0.010000	1.000000	**0.680000**	0.010000	0.040000	0.010000	**0.770000**	**0.770000**
T12	**0.790476**	0.038095	0.028571	0.000000	0.000000	0.009524	0.009524	0.038095	0.009524	0.019048	**0.647619**	1.000000	0.019048	0.028571	0.009524	**0.571429**	**0.571429**
T13	0.050000	0.000000	0.050000	0.000000	0.000000	0.000000	0.000000	0.050000	0.000000	0.000000	0.050000	0.100000	1.000000	0.000000	0.000000	**0.550000**	**0.550000**
T14	0.026316	0.008772	0.000000	0.035088	0.000000	0.000000	0.026316	0.000000	0.000000	0.026316	0.035088	0.026316	0.000000	1.000000	0.035088	**0.280702**	**0.289474**
T15	0.008333	0.016667	0.000000	0.016667	0.008333	0.000000	0.008333	0.000000	0.008333	0.008333	0.008333	0.008333	0.000000	0.033333	1.000000	**0.425000**	**0.425000**
T16	0.017927	0.011432	0.008314	0.010652	0.006755	0.009093	0.016368	0.011691	0.008574	0.012471	0.020005	0.015588	0.002858	0.008314	0.013250	1.000000	**0.983113**
T17	0.017950	0.011707	0.008585	0.011707	0.006504	0.009365	0.017170	0.011707	0.008845	0.012227	0.020031	0.015609	0.002862	0.008585	0.013267	**0.984391**	1.000000

phủ khi không sử dụng 2 kiểm tra này tốt hơn rất nhiều so với khi sử dụng chúng. Đây là một minh chứng cho thấy không phải càng sử dụng nhiều kiểm tra càng tốt.

Hướng nghiên cứu tiếp theo: Bên cạnh bộ kiểm tra NIST SP 800-22 thì còn có rất nhiều các bộ kiểm tra cũng như các kiểm tra thống kê riêng lẻ khác, trong thời gian tới chúng tôi sẽ đánh giá về tính độc lập của các kiểm tra trong bộ kiểm tra của Knuth, là một trong các bộ kiểm tra tính ngẫu nhiên theo thống kê được đưa ra.

Ngoài ra, chúng tôi sẽ phân tích làm rõ độ nhạy của các kiểm tra thống kê nhằm đưa ra các kiểm tra mới cũng như đánh giá "tính ổn định" của từng kiểm tra. Tính ổn định ở đây theo nghĩa giá trị thống kê của kiểm tra đó ít bị ảnh hưởng khi áp dụng kiểm tra đối với các chuỗi đầu vào đã được biến đổi.

Tài Liệu Tham Khảo

[1] Lawrence E Bassham III, Andrew L Rukhin, Juan Soto, James R Nechvatal, Miles E Smid, Elaine B Barker, Stefan D Leigh, Mark Levenson, Mark Vangel, David L Banks, et al. *Sp 800-22 rev. 1a. a statistical test suite for random and pseudorandom number generators for cryptographic applications*. National Institute of Standards & Technology, 2010.

[2] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 2006.

[3] ALİ DOĞANAKSOY, Fatih Sulak, MUHİDDİN UĞUZ, OKAN ŞEKER, and Ziya Akcengiz. Mutual correlation of nist statistical randomness tests and comparison of their sensitivities on transformed sequences. *Turkish Journal of Electrical Engineering & Computer Sciences*, 25(2):655–665, 2017.

[4] A Doğnaksoy, Barış Ege, and Köksal Muş. Extended results for independence and sensitivity of nist randomness tests. In *Information Security and Cryptography Conference, ISC Turkey*, 2008.

[5] Carmina Georgescu, Emil Simion, Alina-Petrescu Nita, and Antonela Toma. A view on nist randomness tests (in) dependence. In *2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pages 1–4. IEEE, 2017.

[6] Peter Hellekalek and Stefan Wegenkittl. Empirical evidence concerning aes. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 13(4):322–333, 2003.

[7] Jorge Augusto Karell-Albo, Carlos Miguel Legón-Pérez, Evaristo José Madarro-Capó, Omar Rojas, and Guillermo Sosa-Gómez. Measuring independence between statistical randomness tests by mutual information. *Entropy*, 22(7):741, 2020.

[8] Onur Koçak. A unified evaluation of statistical randomness tests and experimental analysis of their relations. 2016.

[9] Hoàng Đình Linh, Nguyễn Văn Long. Một tinh chỉnh hiệu quả cho bộ tạo dãy giả ngẫu nhiên massey-rueppel hướng phần cứng. *Journal of Science and Technology on Information security, ISSN 2616-9570*, 06(02):10–17, 2017.

[10] Juan Soto and Juan Soto. *Randomness testing of the advanced encryption standard candidate algorithms*. US Department of Commerce, Technology Administration, 1999.

[11] Fatih Sulak, MUHİDDİN UĞUZ, Onur Kocak, and ALİ DOĞANAKSOY. On the independence of statistical randomness tests included in the nist test suite. *Turkish Journal of Electrical Engineering & Computer Sciences*, 25(5):3673–3683, 2017.

[12] Meltem Sönmez Turan, Ali DoĞanaksoy, and Serdar Boztaş. On independence and sensitivity of statistical randomness tests. In *International Conference on Sequences and Their Applications*, pages 18–29. Springer, 2008.

Sơ lược về tác giả

Đỗ Đại Chí
Đơn vị công tác: Viện Khoa học – Công nghệ mật mã, Ban Cơ yếu Chính phủ
Email: dodaichi2005@gmail.com
Quá trình đào tạo: Tốt nghiệp cử nhân toán tại Đại học Khoa học tự nhiên Hà Nội (2013), Thạc sỹ mật mã tại Đại học Limoges (2019).
Hướng nghiên cứu hiện nay: Lược đồ chữ ký số, An toàn chứng minh được, Mật mã khóa công khai.

Hoàng Đình Linh
Đơn vị công tác: Viện Khoa học – Công nghệ mật mã, Ban Cơ yếu Chính phủ
Email: hoangdinhlinh@bcy.gov.vn
Quá trình đào tạo: Tốt nghiệp cử nhân toán học tại Đại học Khoa học tự nhiên Hà Nội (2014).
Hướng nghiên cứu hiện nay: Bộ sinh số ngẫu nhiên, Tiêu chuẩn đánh giá tính ngẫu nhiên, Mật mã khoá đối xứng.

Trương Minh Phương
Đơn vị công tác: Trường Trung cấp Kỹ thuật mật mã
Email: minhphuongh19@gmail.com
Quá trình đào tạo: Tốt nghiệp Kỹ sư Kỹ thuật mật mã (2012), Thạc sỹ Kỹ thuật mật mã tại Học viện Kỹ thuật mật mã.
Hướng nghiên cứu hiện nay: Hàm băm mật mã, hệ thống bảo mật, hệ thống máy mã.