토픽 모델링 기반 국외 양자내성암호 연구동향 분석

송보연, 김태종*
한국과학기술정보연구원
bysong@kisti.re.kr, *k2boy3@naver.com

Research Trend Analysis of Post-Quantum Cryptography based on Topic Modeling

Boyeon Song and Kim Tae Jong*
Korea Institute of Science and Technology Information (KISTI)

요 약

본 논문은 최근 5년 동안(2018~2022년)의 Web of Science에 등록된 양자내성암호(PQC: Post-Quantum Cryptography) 관련 논문 데이터를 수집하여 토픽 모델링(Topic modeling) 기반으로 어떤 주제가 비중있게 연구되고 있는지 키워드를 중심으로 분석하였다. 수집된 논문을 6개의 핵심 주제로 분류하고 주제별로 주요 키워드를 도출하여 최근 5개년 동안 시계열 추이를 살펴봄으로써 양자내성암호에 관한 연구동향을 파악했다.

Ⅰ. 서 론

양자컴퓨터가 실현되어 사용될 경우 양자컴퓨팅 환경에서도 안전한 암호통신을 하기 위한 미래 암호기술로 양자키분배(QKD: Quantum Key Distribution)와 양자내성암호(PQC; Post-Quantum Cryptography)가 연구 주제로 국내외로 각광을 받고 있다[1]. 양자내성암호란 양자컴퓨터가 도래한 이후의 암호기술이라는 의미로 양자컴퓨팅으로 가능한 모든 공격에 대해 안전한 내성을 가진 공개키 암호를 말한다.[2]

본 논문에서는 최근 5년 동안(2018~2022년)의 Web of Science[3]에 등록된 양자내성암호 관련 논문 데이터를 수집하여 토픽 모델링(Topic Modeling)[4,5] 기반으로 연구 주제에 대한 동향을 분석하고자 한다.

Ⅱ. 양자내성암호 연구동향 분석

본론에서는 양자내성암호에 대한 최근 주요 연구 주제를 분석하기 위한 논문 데이터 수집 및 분석 방법에 대해서 소개하고, 토픽 모델링 기반으로 논문 데이터를 분석한 결과를 핵심 주제별로 제시한다.

2.1. 연구방법

2.1.1. 데이터 수집

분석대상 논문은 2023년 1월 1일을 기준으로 최근 5년 동안 Web of Science에 등록된 양자내성암호 관련 논문이다. 데이터 수집을 위한 검색어는 'post-quantum cryptography'이며, 검색 범위는 논문 제목, 키워드, 초록으로 설정하여 검색했다. 검색 결과 총 624건의 논문이 수집되었으며, 이 중 중복되거나 연구목적에 적합하지 않은 논문 19건을 제외하여 총 605건의 논문을 대상으로 분석을 수행했다. 연도별 논문현황은 [그림 1]과 같다.

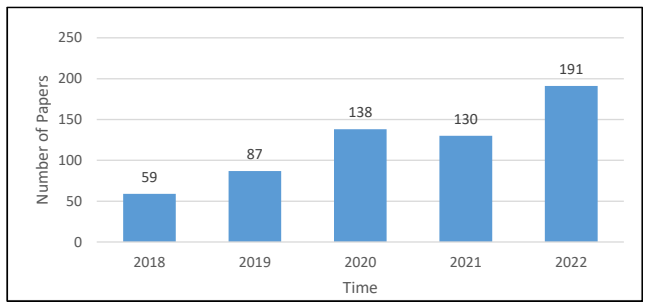

그림 1. 연도별 변화
Fig. 1. Changes by year

2.1.2. 분석 방법

본 연구의 목적은 국외 양자내성암호 관련 연구에 나타난 주요 키워드와 토픽이 무엇인지 파악하여, 양자내성암호 연구 동향을 분석하는 것이므로, 텍스트 등의 비정형 데이터에서 핵심 주제를 도출하는 데에 유용하게 사용되고 있는 토픽 모델링 분석 방법을 적용했다. 토픽 모델링 분석 방법은 텍스트 데이터를 대상으로 머신러닝 알고리즘을 활용하여 키워드와 토픽을 자동으로 추출하는 방법으로서, 토픽 모델링 분석 방법 중 LDA 기법은 문서 집합에서 핵심 키워드를 출현확률(Probability)에 따라 토픽별로 분류하는 분석 기법으로, 연구 동향을 파악하는 연구에서 널리 활용되고 있다[4].

분석 프로그램으로는 국산 프로그램인 NetMiner 4.5를 사용했다. 최적화된 토픽 수를 결정하기 위해, 토픽 수를 4부터 13까지 설정하고 α값을 0.01, 0.05, 0.1로 설정하여 실루엣 계수(Silhouette coefficient)를 측정했다(Iteration: 1,000). 실루엣 계수를 활용한 검증방법은 Panichella[5] 등 연구를 통해 제안한 검증방법으로서, 특정 데이터가 소속된 클러스터의 데이터와 얼마나 가까이 클러스터링이 되었으며, 다른 클러스터의 데이터와 얼마나 멀리 떨어져 있는지 나타내는 지표다.

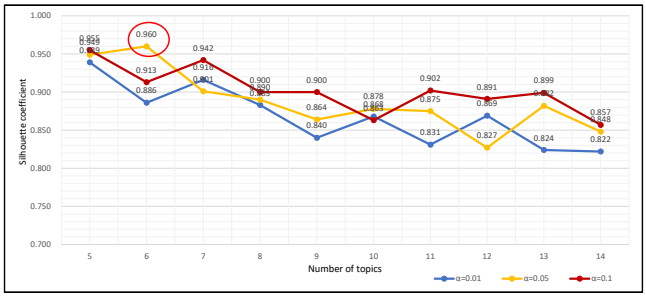

그림 2. 토픽 모델링 실루엣계수 측정 결과
Fig. 2. Topic modeling silhouette coefficient measurement result

실루엣 계수는 토픽 내부의 유사성과 토픽 간의 차별성을 계량적으로 측정할 수 있는 지수로서, -1에서 +1까지의 범위를 가지며, +1에 가까울수록 토픽 모델링이 최적화된 것을 뜻한다[2]. 실루엣 계수를 측정한 결과, [그림 2]와 같이 α값이 0.05, 토픽 수가 6개일 경우 실루엣 계수가 0.960으로서 +1에 가장 가깝게 나타났으므로, 해당 값을 기준으로 토픽 모델링 분석을 수행했다.

2.2 연구결과

양자내성암호 관련 국외 논문 605건을 대상으로 토픽 모델링 분석 방법을 통해 토픽을 도출한 결과 [그림 3]과 같이 6개의 토픽과 각 토픽별 비중이 나타났으며, 시계별 비중 변화는 [그림 4]와 같다.

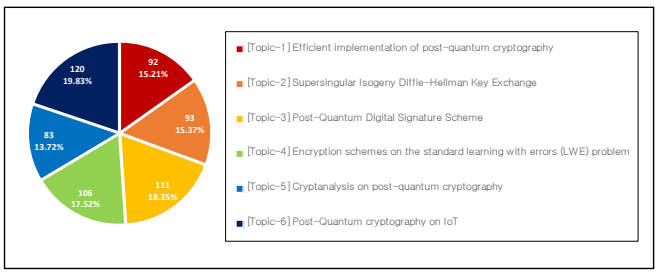

그림 3. 토픽별 비중
Fig. 3. Proportion by topic

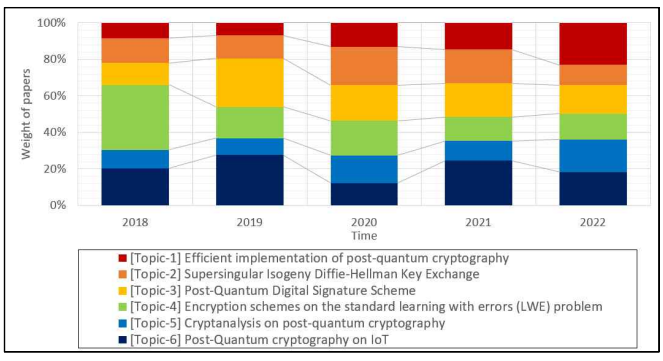

그림 4. 연도별 토픽 변화
Fig. 4. Topic change by year

2.2.1. [Topic-1] 분석 결과

[Topic-1]에서는 'polynomials', 'multiplication', 'NIST', 'NTT', 'standardization', 'matrix', 'GPU', 'ring', 'encryption', 'accelerator' 등이 주요 키워드로 도출됨에 따라 관련 논문을 검토하여, 토픽명을 '양자내성암호의 효율적 실행(Efficient implementation of post-quantum cryptography)'으로 정의했으며, 시계열 변화는 증가하는 것으로 나타났다.

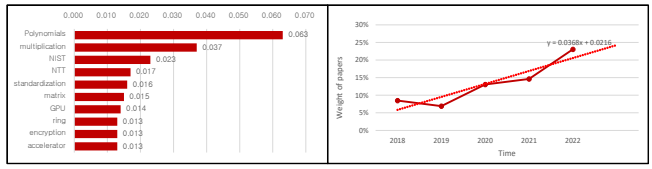

그림 5. [Topic-1] 주요 키워드 및 시계열 변화
Fig. 5. [Topic-1] Keywords and Time Series Changes

2.2.2. [Topic-2] 분석 결과

[Topic-2]에서는 'SIDH(SIKE)', 'isogeny', 'FPGA', 'NIST', 'multiplier', 'side-channel', 'sampler', 'Xilinx', 'CSIDH', 'key encapsulation' 등이 주요 키워드로 도출됨에 따라 관련 논문을 검토하여, 토픽명을 '초특이성 아이소제니 디피-헬만 키 교환(Supersingular Isogeny Diffie-Hellman Key Exchange)'으로 정의했으며, 시계열 변화는 미세하게 증가된 것으로 나타났다.

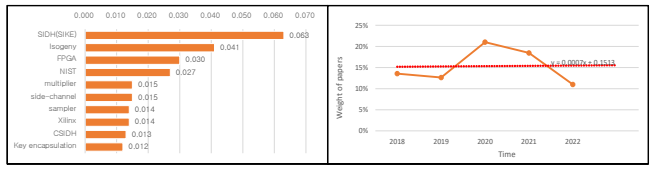

그림 5. [Topic-2] 주요 키워드 및 시계열 변화
Fig. 5. [Topic-2] Keywords and Time Series Changes

2.2.3. [Topic-3] 분석 결과

[Topic-3]에서는 'signature', 'blockchain', 'lattice', 'random oracle', 'ring signature', 'digital signature', 'verification', 'identification', 'authentication', 'certificate' 등이 주요 키워드로 도출됨에 따라 관련 논문을 검토하여, 토픽명을 '양자내성 서명 기법(Post-Quantum Digital Signature Scheme)'으로 정의했으며, 시계열 변화는 감소하는 것으로 나타났다.

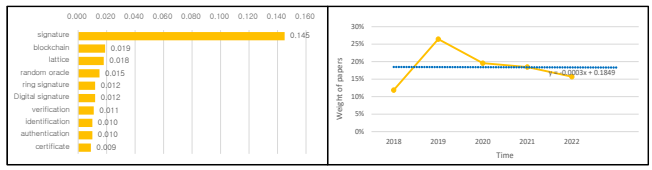

그림 5. [Topic-3] 주요 키워드 및 시계열 변화
Fig. 5. [Topic-3] Keywords and Time Series Changes

2.2.4. [Topic-4] 분석 결과

[Topic-4]에서는 'encryption', 'LWE', 'lattice', 'McEliece', 'ciphertext', 'decoding', 'complexity', 'reduction', 'cloud', 'vector' 등이 주요 키워드로 도출됨에 따라 관련 논문을 검토한 결과, 토픽명을 '선형잡음문제(LWE) 기반의 암호 기법(Encryption schemes on the standard learning with errors (LWE) problem)'으로 정의했으며, 시계열 변화는 감소하는 것으로 나타났다.

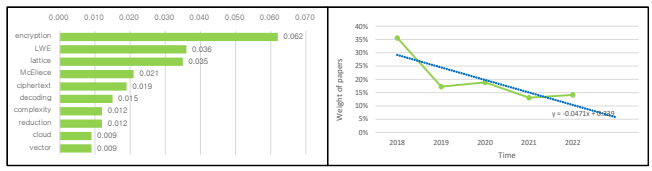

그림 5. [Topic-4] 주요 키워드 및 시계열 변화
Fig. 5. [Topic-4] Keywords and Time Series Changes

2.2.5. [Topic-5] 분석 결과

[Topic-5]에서는 'encryption', 'complexity', 'NIST', 'leakage', 'side-channel', 'query', 'standardization', 'KEM', 'cryptanalysis', 'transformation' 등이 주요 키워드로 도출됨에 따라 관련 논문을 검토하여, 토픽명을 '양자내성암호에 대한 암호학적 분석(Cryptanalysis on post-quantum cryptography)'으로 정의했으며, 시계열 변화는 증가하는 것으로 나타났다.

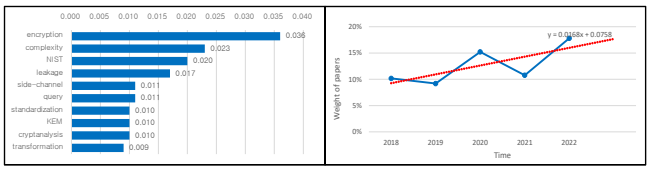

그림 5. [Topic-5] 주요 키워드 및 시계열 변화
Fig. 5. [Topic-5] Keywords and Time Series Changes

2.2.6. [Topic-6] 분석 결과

[Topic-6]에서는 'IoT', 'QKD', 'authentication', 'privacy', 'secret key', 'noise', 'vehicle', 'eavesdropper', 'entanglement', 'cloud' 등이 주요 키워드로 도출됨에 따라 관련 논문을 검토하여, 토픽명을 'IoT에 적용 가능한 양자내성암호(Post-Quantum cryptography on IoT)'로 정의했으며, 시계열 변화는 감소하는 것으로 나타났다.

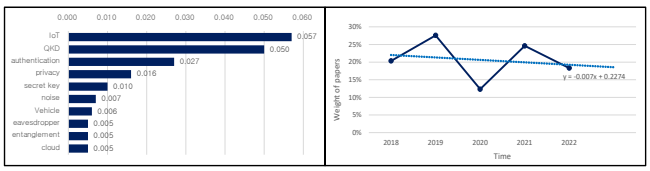

그림 5. [Topic-6] 주요 키워드 및 시계열 변화
Fig. 5. [Topic-6] Keywords and Time Series Changes

Ⅲ. 결론

Web of Science에 등록된 최근 5년간(2018~2022년)의 논문을 바탕으로 토픽 모델링을 적용하여 주요 주제별 키워드 도출 및 시계열 연구 동향 변화를 분석해보았다. 양자컴퓨터의 도래를 대비하기 위해 양자내성암호에 대한 연구는 18년 이후 현재까지 크게 증가하고 있고, 주제의 특성 상 앞으로 더욱 증가할 것으로 예상된다. 수집된 논문 데이터 분석으로 도출된 6개의 핵심 주제에 대한 추이를 살펴봤을 때 초기에는 특정한 주제에 대한 연구가 주를 이루었다면 점차적으로 6개의 핵심 주제로 연구가 고르게 분산되어 진행되는 것을 확인할 수 있었다. 본 논문에서 제시한 주요 6개의 주제에 대한 연구 뿐 아니라 다각적인 측면에서 양자내성암호에 관한 연구가 활발하게 확대 및 진행될 것으로 전망된다.

ACKNOWLEDGMENT

본 연구는 2023년도 한국과학기술정보연구원(KISTI) 기본사업 과제로 수행한 것입니다.

참 고 문 헌

[1] https://ko.wikipedia.org/wiki/양자_후_암호

[2] http://wiki.hash.kr/index.php/양자내성암호

[3] https://www.webofscience.com, Web of Science, Clarivate

[4] D. M. Blei, "Probabilistic topic models", Communication of the ACM, 55(4), pp. 77-84, April 2012

[5] A. Panichella, B. Dit, R. Oliveto, M. Di Penta, D. Poshynanyk, and A. De Lucia, "How to effectively use topic models for software engineering tasks? An approach based on Genetic Algorithms", 2013 35th International Conference on Software Engineering(ICSE), San Francisco, CA , pp. 522-531, May 2013 (https://doi.org/10.1109/ICSE.2013.6606598)