Chapter 39
Scientometric and Wikipedia Pageview Analysis

Alexander Glavackij, Sarah Ismail, and Dimitri Percia David

39.1 Introduction

This chapter explores trends in data protection and encryption technologies across different technologies. The technologies analyzed are taken from the previous chapters.

Any trend assessment concerning data protection and encryption technologies constitutes a challenging task for various reasons. The swift development of the security technologies brings a myriad of novel protocols, tools, and procedures, whose technological readiness levels (TRL) also evolve rapidly [1]. Also, while some technologies thrive, others stagnate or vanish in favour of more market-adapted technologies or enhanced operational implementation [2]. Moreover, in such a fast-paced and growing environment, opportunities and threats evolve quickly, making it difficult to evaluate the whole spectrum of technologies available on the market [3]. Consequently, evaluations of the security consequences of the arrival and evolution of such technologies on data protection are complex.

Following the previous individual analysis of the data protection and encryption technologies, we evaluate these technologies through time by benchmarking a development indicator—the *attention* paid by different communities [4].

The original version of this chapter has been revised. Third author name has been corrected. The correction to this chapter can be found at https://doi.org/10.1007/978-3-031-33386-6_42

A. Glavackij · S. Ismail
Cyber-Defence Campus, Thun, Switzerland
e-mail: alexander.glavackij@ar.admin.ch; sarah.ismail@ar.admin.ch

D. Percia David (✉)
HES-SO Valais-Wallis, Sion, Switzerland
e-mail: dimitri.perciadavid@hevs.ch

V. Mulder et al. (eds.), *Trends in Data Protection and Encryption Technologies*,
https://doi.org/10.1007/978-3-031-33386-6_39

39.2 Analysis

39.2.1 Scientometric analysis

We conduct a scientometric analysis of the book's technologies to better understand how they evolved over the last 20 years. Most cryptographic technologies are the result of long-term research efforts. Therefore, we analyze the number of associated scientific works through time for each technology, which can be seen as an indicator of scientific interest in that technology [5]. Growing attention points toward promising or emerging technologies, as researchers tend to dedicate significant resources to potentially valuable technologies. Conversely, low interest in a given technology correlates with the lack of technological novelty and obsolescence.

To provide such a scientometric analysis, we use the OpenAlex dataset, which describes scholarly entities (works, authors, institutions, venues, and concepts) and their connectivity patterns using a graph structure.[1] Importantly, each scholarly work has concepts associated with it that are represented in the paper. OpenAlex organizes publications' concepts into a tree structure, where general concepts are parents of more fine-grained ones. OpenAlex has 65,026 concepts, ranging from Political Science to Physics. Scientific works are tagged automatically using a classification model trained on the Microsoft Academic Graph (MAG) [6]. Thus, OpenAlex provides a taxonomy of topics discussed in the scientific literature, used here to retrieve scientific works tagged with this book's 38 technologies. We scrape the scientific works tagged with those 38 technologies, taking for each a monthly count of the number of published papers. This yields a time series for each technology, which we use to analyze the technologies over time.

The technologies' time series display different development patterns; therefore, we cluster them according to the exhibited pattern into three classes: no growth, moderate growth, and strong growth. We calculate the clusters in the following manner: we divide the average number of publications during the first 3 months in 2022 by the average number of publications during the first 3 months in 2012. We refer to the resulting ratio as *growth ratio*. If growth ratio < 1.05, we deem the technology as not growing. The technology exhibits moderate growth if $1.05 <$ growth ratio < 2. The technology thrives if growth ratio > 2.

Additionally, we cluster the technologies into low, moderate, and high-interest technologies. A technology is a high-interest technology if the average monthly publication count is $c \geq 50$, a moderate-interest technology if $15 \leq c < 50$, and a low-interest technology if $c < 15$. The growth pattern and interest level form a two-dimensional matrix where we can arrange the technologies. Table 39.1 shows the resulting matrix for 24 selected technologies.

We emphasize interesting patterns. High-interest technologies which have been researched extensively, include Blockchain, Hash Function, and Asymmetric

[1] https://docs.openalex.org/.

Table 39.1 Selected technologies of this book assorted into a two-dimensional matrix, created by clustering the technologies by their past growth and interest in the research community

Growth Pattern \ Interest	Low Interest	Moderate Interest	High Interest
No Growth	Confidential Computing, Digital Rights Management, Disk Encryption	Authentication, Digital Signature, Identity Management, Key Management	Asymmetric Encryption
Moderate Growth	Electronic Voting, Functional Encryption	Quantum Cryptography, Random Number Generation, Symmetric Cryptography	Biometrics, Hash Function
Strong Growth	Hardware Acceleration, Hardware Security Module, Post-Quantum Cryptography, Zero-Knowledge Proof	Differential Privacy, Homomorphic Encryption, Quantum Key Distribution	Blockchain

Encryption in this cluster. Except for Blockchain, these technologies represent the backbone of today's cybersecurity landscape. However, Blockchain is the only one exhibiting strong growth. This might indicate that Blockchain technology has a large part of its development ahead of it. Moderate-interest technologies represent more specialized techniques and methodologies that have established themselves. Digital Signatures, Authentication, and Key Management are well-known and widely used technologies, but interest in them is not growing further, indicating technical convergence. Emerging technologies, especially Differential Privacy and Quantum-related technologies, exhibit growth and can be counted on to become more critical in the future. Low interest and not growing technologies are niche technologies, like Disk Encryption and Functional Encryption. However, some low-interest technologies exhibit strong growth, like Post-Quantum Cryptography, Zero-Knowledge Proof, and Hardware Security Modules. These technologies have been relatively recently established, and the research interest indicates that most discoveries in those technologies are yet to be made (Fig. 39.1).

39.2.2 Evolution of public attention

To explore more, we look at the evolution over time of public attention to technologies through Wikipedia's pageviews statistics. The motivation of the public to know more about a technology provides good information on the position and the popularity of the technologies [7]. Wikipedia's pageviews statistics show the number of pages visited over a given period at a given chosen frequency (either daily, monthly, or yearly data—in this work, we use the monthly frequency). Such statistics cover each page. The statistics do not consider the time Internet users spend on a page. Whatever its duration, it will be counted as a view. We collect data for 37 technologies over 82 months, from July 2015 to April 2022. Again, we group the technology time series into three classes: no growth, moderate growth, and high growth. Clusters are calculated as follows: we divide the average pageviews

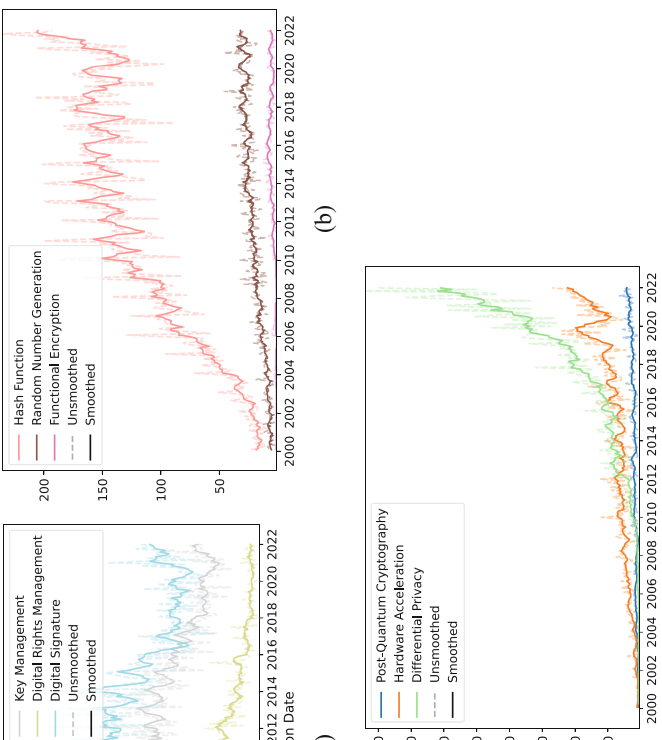

Fig. 39.1 Number of monthly scientific works established between 2000 and 2022, taken from OpenAlex, for three samples from each growth pattern cluster. The No Growth technologies in (a) have already achieved their peak of maximal interest, i.e., interest in those technologies is waning. For the Moderate Growth technologies (b), growth has been slowing recently. If researchers do not discover new research areas in those fields, interest will decrease further, and these technologies will shrink soon. The fast-growing technologies (c) are fairly recently established technologies and can be expected to be further developed. (**a**) No growth. (**b**) Moderate growth. (**c**) Strong growth

Table 39.2 Technologies of this book assorted to a two-dimensional matrix created by clustering the technologies by their past growth and public interest

Growth Pattern \ Interest	Low Interest	Moderate Interest	High Interest
No Growth	Authentication, Confidential Computing, Disk Encryption, Electronic Voting, Email Security, Hardware Security Module, Identity Management, Key Management, Quantum Cryptography, Secure Messaging, Secure Operating System, Symmetric Cryptography, Tunneling	Biometrics, Digital Rights Management, Digital Signature	Asymmetric Encryption, Hash Function
Moderate Growth	Differential Privacy, Functional Encryption, Hardware acceleration, Homomorphic Encryption, Quantum Key Distribution	Random Number Generation	
Strong Growth	Identity-based Cryptography, Multi-party Threshold Cryptography, Post-quantum Cryptography, Private Set Intersection, Searchable Symmetric Encryption, Secure Multi-Party Computation, Trusted Execution Environment, Zero-knowledge Proof		Blockchain

of the last 3 months by the average pageviews of the last 3 months from 6 years ago. We refer to the resulting ratio as *growth ratio*. If growth ratio < 1.05, we deem the technology as not growing. The technology exhibits moderate growth if $1.05 <$ growth ratio < 2. The technology thrives if $2 <$growth ratio. We also cluster the technologies into low, moderate, and high-interest technologies. A technology is a high-interest technology if the average number of pageviews per month is $c \geq 50{,}000$, a moderate-interest technology if $25{,}000 \leq c < 50{,}000$, and a low-interest technology if $c < 25{,}000$. We provide the two-dimensional matrix clustering the technologies according to their growth in Table 39.2 (Fig. 39.2).

Again, the technologies attracting significant public interest are Blockchain, Hash Function, and Asymmetric Encryption. Blockchain shows strong growth, unlike Hash Function and Asymmetric Encryption, which show no growth. Again, technologies with more specialized techniques and methodologies, such as Digital Signatures and Biometrics, are seeing moderate interest. Low-interest and no-growth technologies are niche technologies, such as Disk Encryption, or long-standing technologies, such as Email Security. However, some low-interest technologies, such as Post-quantum Cryptography, still show strong growth (Fig. 39.3).

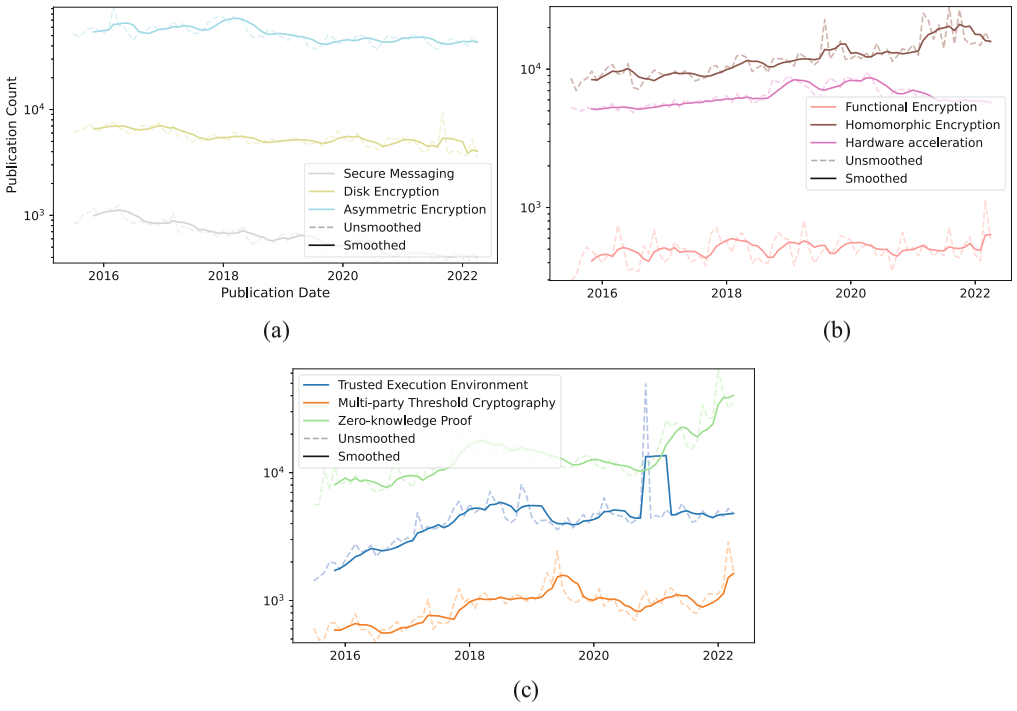

Fig. 39.2 Number of monthly pageviews of Wikipedia between 2015 and 2022 for three samples from each growth pattern cluster. The No Growth technologies in (a) are already at their peak of maximal interest, i.e., interest in those technologies is waning. For the Moderate Growth technologies (b), growth has been slowing recently. If researchers do not discover new research areas in those fields, interest will decrease further, and these technologies are expected to shrink soon. The fast-growing technologies (c) are fairly recently established technologies and can be expected to be further developed. (**a**) No growth. (**b**) Moderate growth. (**c**) Strong growth

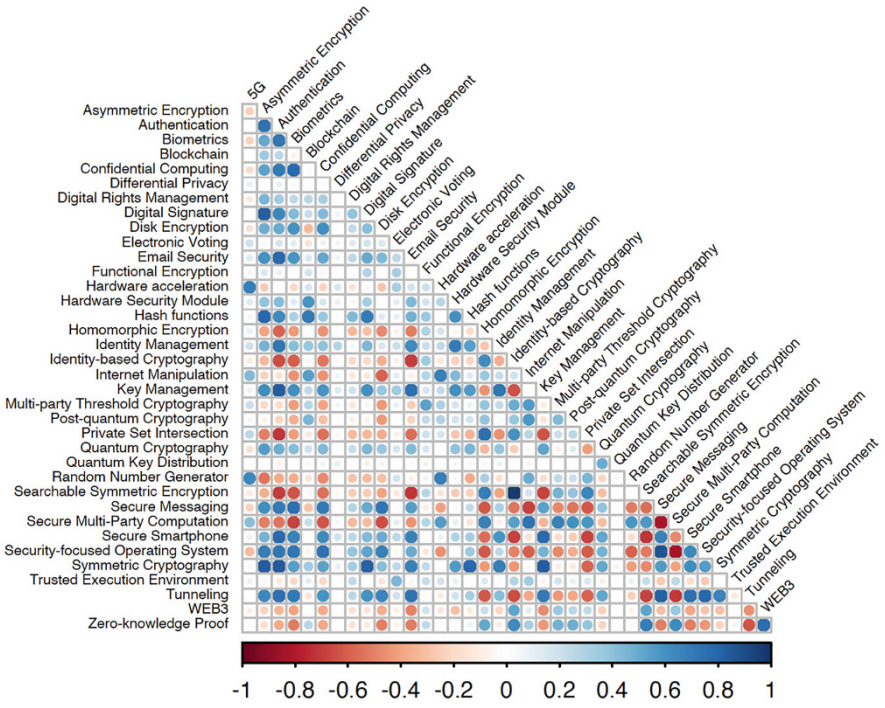

Fig. 39.3 Correlation between pageviews of technologies in Wikipedia from July 2015 to April 2022

39.2.3 Correlation Analysis

In order to proceed to an exploratory data analysis and get an idea of the potential existing relationships between these technologies, we display a correlation matrix of Wikipedia's monthly pageviews. However, these correlations can potentially contain confounding factors and spurious relationships (as time series are not stationary). Figure 39.4 shows positive or negative correlations between page views. For instance, we notice that "Identity-based encryption" and "Searchable symmetric encryption" highly correlate. On the other hand, "Quantum Key Distribution" has no or a very weak correlation with all the other technologies.

39.2.4 Comparison of public and expert attention

The relationship between these proxies for the two types of attention diverges over time. However, graphically, we observe that expert attention follows public attention by a few months. For instance, in the case of Random Number generation, public

Fig. 39.4 These plots compare public attention from Wikipedia page views (blue line) and expert attention from the number of publications on OpenAlex (red line). We discard outliers, and the study period is from July 2015 to April 2022. Data is provided on a monthly frequency. The method used on the data is the robust z-score with a scale from −5 to 5 to have a better view and eliminate large outliers

attention increased around 2019, while expert attention started to pick up 1 year later.

39.3 Conclusion

In conclusion, this chapter evaluates data protection and encryption technology trends through time. We used a benchmarking development indicator, the attention brought by different communities, to perform the analysis. This attention was measured through a scientometric analysis of the production of scientific works and the public attention was given to these technologies through Wikipedia pageviews. Our results showed that high-interest technologies like Blockchain, Hash Function, and Asymmetric Encryption are widely researched and used, but only Blockchain exhibited strong growth. Moderate-interest technologies like Digital Signatures, Authentication, and Key Management have established themselves but need to show growth, indicating technical convergence. Finally, emerging technologies like Differential Privacy and Quantum-related technologies showed growth, indicating their potential to become more critical in the future. This analysis provides valuable insights into the development of data protection and encryption technologies and their impact on the security landscape.

References

1. Yu-Wei Chang and Jiahe Chen. What motivates customers to shop in smart shops? the impacts of smart technology and technology readiness. *Journal of Retailing and Consumer Services*, 58:102325, 2021.
2. Bram Faber. A tale of three technologies: A survival analysis of municipal adoption of websites, twitter, and youtube. *Digital Government: Research and Practice*, 2022.
3. Alessandro Merendino, Sally Dibb, Maureen Meadows, Lee Quinn, David Wilson, Lyndon Simkin, and Ana Canhoto. Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research*, 93:67–78, 2018.
4. JooYoung Lee, Siqi Wu, Ali Mert Ertugrul, Yu-Ru Lin, and Lexing Xie. Whose advantage? measuring attention dynamics across youtube and twitter on controversial topics. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 573–583, 2022.
5. Tugrul U Daim. *Digital Transformation: Evaluating Emerging Technologies*, volume 6. World Scientific, 2020.
6. Zhihong Shen, Hao Ma, and Kuansan Wang. A web-scale system for scientific knowledge exploration.
7. Yujia Yang, Shi Lu, Huan Zhao, and Xiaoqian Ju. Predicting monthly pageview of wikipedia pages by neighbor pages. In *Proceedings of the 2020 3rd International Conference on Big Data Technologies*, pages 112–115, 2020.