Computer Science Review 42 (2021) 100431

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

# Research communities in cyber security: A comprehensive literature review

Sotirios Katsikeas *, Pontus Johnson, Mathias Ekstedt, Robert Lagerström

*KTH Royal Institute of Technology, Division of Network and Systems Engineering, Teknikringen 33, floor 3, SE 100 44 Stockholm, Sweden*

ARTICLE INFO

*Article history:*
Received 26 August 2020
Received in revised form 5 March 2021
Accepted 25 August 2021
Available online 11 September 2021

*Keywords:*
Security
Clustering
Community
Systematic literature review

ABSTRACT

In order to provide a coherent overview of cyber security research, the Scopus academic abstract and citation database was mined to create a citation graph of 98,373 authors active in the field between 1949 and early 2020. The Louvain community detection algorithm was applied to the graph in order to identify existing research communities. The analysis discovered twelve top-level communities: access control, authentication, biometrics, cryptography (I & II), cyber–physical systems, information hiding, intrusion detection, malwares, quantum cryptography, sensor networks, and usable security. These top-level communities were in turn composed of a total of 80 sub-communities. The analysis results are presented for each community in descriptive text, sub-community graphs, and tables with, for example, the most-cited papers and authors. A comparison between the detected communities and topical areas defined by other related work, is also presented, demonstrating a greater researcher emphasis on cryptography, quantum cryptography, information hiding and biometrics, at the expense of laws and regulation, risk management and governance, and security software lifecycle.



**Contents**

1. Introduction ... 2
2. Method ... 2
   2.1. Scopus as data source ... 2
   2.2. Data retrieval and processing ... 2
   2.3. Search criteria ... 2
      2.3.1. Initial scoping by keywords ... 2
      2.3.2. Fine-tuning the keywords ... 2
      2.3.3. Collecting 59,782 articles ... 3
   2.4. Collected meta-data ... 3
   2.5. Producing the author graph ... 3
   2.6. Community detection ... 3
   2.7. Community graph ... 4
   2.8. Sub-community detection and description ... 5
3. Data ... 5
4. Results & analysis ... 5
   4.1. Cryptography I & II ... 6
   4.2. Sensor networks ... 7
   4.3. Information hiding ... 8
   4.4. Intrusion detection ... 10
   4.5. Malwares ... 11
   4.6. Biometrics ... 12
   4.7. Cyber–physical systems ... 13
   4.8. Authentication ... 15
   4.9. Usable security ... 16
   4.10. Access control ... 16

* Corresponding author.
*E-mail addresses:* sotkat@kth.se (S. Katsikeas), pontusj@kth.se (P. Johnson), mekstedt@kth.se (M. Ekstedt), robetl@kth.se (R. Lagerström).

https://doi.org/10.1016/j.cosrev.2021.100431

4.11. Quantum cryptography ..... 18
5. Related work ..... 19
5.1. Comparison to CyBoK ..... 20
5.2. Comparison to Baset and Denning ..... 21
6. Discussion ..... 21
7. Summary ..... 22
Declaration of competing interest ..... 22
References ..... 22

## 1. Introduction

The cyber security research community is an eclectic group, addressing a diverse set of research questions, based on multifarious theories and deploying sundry methods, making it difficult to obtain a comprehensive grasp of this league. Using quantitative methods, the present work aims to summarize the activities of this group of researchers in a coherent manner. In a citation graph of 98,373 authors active in the field of cyber security between 1949 and early 2020, we identify twelve distinct communities focusing on various topics, such as Malware, Usable Security, Intrusion Detection and Access Control. Each community is described e.g. in terms of research foci, publication fora, and sub-community evolution. Ever since Thomas Kuhn's seminal work *The structure of scientific revolutions* [1], philosophers of science have been aware of the impact of social organization on the scientific endeavor. It is therefore not surprising to discover that cyber security research communities and sub-communities are not solely explained by their topical foci, but sometimes by other factors, such as geography.

Section 2 details the methods used to collect and analyze the abstract and citation data on which the article is based. In Section 3, an overview of the collected data is done and some metadata are presented. Section 4 contains the results of the analysis, presenting in some detail each of the twelve research communities. This is followed by related works, including a comparison with other attempts to summarize the field. Section 6 consists of a discussion of the results, considering validity and reliability. The article is concluded with a summary in Section 7.

## 2. Method

### 2.1. Scopus as data source

Scopus was selected as our data source, first, because of its comprehensive content, since it contains more than 25,000 active titles from 7000 publishers that are reviewed and selected by an independent review board.[1] According to the official Scopus website, "*over the past 3 years, Scopus has added over 195 million more cited references, dating back to 1970, to complement the database's existing records that date back to 1788 and further increase the depth of content*".[2] Compared with other academic publication databases, such as IEEE Xplore and ACM Digital Library, Scopus is one of the most comprehensive [2] and includes data from both IEEE and ACM publications (among others). Scopus has almost the same database size as Web of Science but offers a more flexible and powerful open-access API.

### 2.2. Data retrieval and processing

The Scopus API was used in a Python script to search and collect meta-information about a set of articles that were obtained using a custom search query. This meta-information was then pre-processed and stored on Google Cloud Datastore for post-processing and analysis. After the data retrieval phase, which required approximately two weeks, the data could easily and promptly be retrieved from Datastore in under 4 min, so that the analysis could be performed. In the analysis, the main author graph was generated and then partitioned, as will be described later, and then the information and graphs for each of the communities obtained were calculated and provided to the user.

### 2.3. Search criteria

As mentioned in the introduction, our goal is to analyze the cyber security research community. Accordingly, any published article related to either `information security` or `cyber security` should be initially considered to be of interest. These two keywords were the starting search criteria in our study because they are usually (albeit inaccurately) used as synonyms in research. These two terms were expected to capture a sufficient number of articles to allow us to collect the relevant keywords for the next phase.

#### 2.3.1. Initial scoping by keywords

Initially, articles containing the keywords `information security` or `cyber security` in their title, abstract, or keywords were searched for, and 27,654 documents were retrieved. The keywords most commonly used in these articles were subsequently ordered in descending frequency. Among these, those deemed excessively generic to fit in the domain of information and cyber security (e.g., `Internet`), were discarded, resulting in the following set of top 21 keywords.

| | | |
|---|---|---|
| Access Control | Authentication | Computer Crime |
| Computer Security | Computer Viruses | Cryptography |
| Cyber Security | Cyber-attacks | Cybersecurity |
| Digital Watermarking | Information Security | Intrusion Detection |
| Malware | Mobile Security | Network Security |
| Privacy | Security Of Data | Security Policy |
| Security Requirements | Security Systems | Steganography |

#### 2.3.2. Fine-tuning the keywords

The top 25 articles for each of these keywords were manually examined to detect possible outliers. If such articles were detected, the search query was fine-tuned to exclude them. The results of this examination are presented below.

The `Access Control` keyword resulted in a large number of articles that were not security-related; these were primarily concerned with media or physical access control. To handle this, the `AND TITLE-ABS-KEY (''Security'')` filter was applied on the `Access Control` term to discard these articles. For validation, a search for `(KEY (''Access Control'') AND NOT TITLE-ABS-KEY (''Security''))` was also performed, resulting in 17,778 articles, the great majority of which were not security-related. The same approach was followed for the `Privacy` keyword.

It should also be noted that Scopus handles the keywords provided inside double quotation marks in a particular manner

[1] https://www.elsevier.com/solutions/scopus/how-scopus-works
[2] https://www.elsevier.com/solutions/scopus/how-scopus-works/content

**Fig. 1.** Author graph of the research communities in cyber security. Each node represents an author, colors denote research communities and the size of each node is proportional to the number of citations.

(called "search for a loose or approximate phrase") whereby issues related to capitalization, singular, plural, or hyphenated words are eliminated. Furthermore, multi-word keywords such as `Data Privacy` are covered by simpler ones such as `Privacy`. To ensure that this is indeed the case, we performed two searches for these two keywords, and we concluded that the results of `KEY (''Privacy'')` include all the results of `KEY (''Data privacy'')`.

When we examined the `Mobile Security` keyword, we were surprised by the fact that there were 577 results related to "cytology". To eliminate them from the results of the main query, we applied the `AND NOT KEY (''Cytology'')` filter on the keyword term. We opted to exclude them at scraping time to reduce the time required to filter them out during the subsequent analysis.

Finally, `AND TITLE-ABS-KEY (''Security'')` was also applied on `Digital Watermarking` because 5832 results of that term were related to watermarking but not for security purposes.

#### 2.3.3. Collecting 59,782 articles

The top 21 relevant keywords together with the improvements explained above were used in a logical disjunction (`security of data` OR `network security` OR ...) as the core of the main search in the Scopus database in a search query that was also restricted (in terms of subject area) to computer science, engineering, social sciences, decision sciences, multidisciplinary, or undefined (to exclude articles clearly off topic), and restricted to the English language. The full query that was used can be found at.[3]

This search query resulted in 320,907 articles. Unfortunately, owing to the search quota limitations of the Scopus API, we were forced to limit the results to the top 5000 most cited articles for large queries. Thus, we performed a distinct search for each year. To avoid under-representing the years with a large number of articles (e.g., selecting all 4374 articles from 2001, but only approximately 19% of the 26,642 articles produced in 2016), we selected the same fraction of articles for each year, with the peak of the 5000 most-cited articles from the peak year of 2019, when 33,884 articles were produced. Hence, for each year, the 19.9% most-cited articles were collected. In total, this resulted in a dataset of 59,782 articles.

### 2.4. Collected meta-data

For each article, the following meta-data were collected: (i) EID (unique academic work identifier in Scopus), (ii) authors, (iii) title, (iv) source, (v) keywords, (vi) publication date, and (vii) references.

For each author, the following information was gathered: (i) Scopus author ID, (ii) surname, (iii) given name, and (iv) affiliation.

Finally, for each affiliation, the following information was gathered: (i) Scopus affiliation ID, (ii) name, and (iii) country.

### 2.5. Producing the author graph

Based on the collected data, a citation graph was generated, in which all authors are linked to each other according to citations. In the graph, authors are represented by nodes, and undirected edges between nodes indicate that at least one author has cited the other at least once, and the size of the nodes is related to the number of citations each author has. The main author graph is shown in Fig. 1 (to reduce its size, the graph only contains the authors that have more than 12 citations globally, and the edges are hidden).

### 2.6. Community detection

The author graph is a social graph, in the sense that it represents relations between people. A significant amount of research on the analysis of such graphs has been conducted, particularly on community detection. One of the best-performing algorithms for community detection in large graphs is the Louvain method, proposed by Blondel et al. [3]. As the authors write,

"*The problem of community detection requires the partition of a network into communities of densely connected nodes, with the*

[3] https://git.io/JvRam

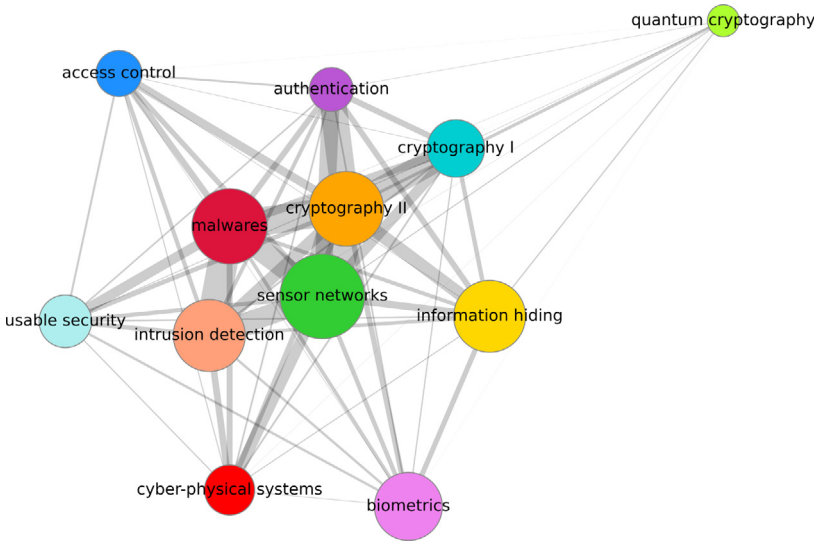


**Fig. 2.** The security research community graph generated from the analyzed data.

*nodes belonging to different communities being only sparsely connected. Precise formulations of this optimization problem are known to be computationally intractable. Several algorithms have therefore been proposed to find reasonably good partitions in a reasonably fast way.*"

The algorithm aims to find a graph partition that maximizes modularity, which is a scalar value between −1 and 1 that "measures the density of links inside communities as compared to links between communities", defined as follows:

$$Q = \frac{1}{2m}\sum_{i,j}[A_{ij} - \frac{k_i k_j}{2m}]\delta(c_i, c_j)$$

where $A_{ij}$ represents the weight of the edge between vertices $i$ and $j$, $k_i = \sum_j A_{ij}$ is the sum of the weights of the edges attached to vertex $i$, $c_i$ is the community to which vertex $i$ is assigned, the $\delta$-function $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise, and $m = \frac{1}{2}\sum_{i,j} A_{ij}$.

The Louvain community detection algorithm operates as follows (in the words of the authors):

> Our algorithm is divided in two phases that are repeated iteratively. Assume that we start with a weighted network of N nodes. First, we assign a different community to each node of the network. So, in this initial partition there are as many communities as there are nodes. Then, for each node i we consider the neighbors $j$ of $i$ and we evaluate the gain of modularity that would take place by removing $i$ from its community and by placing it in the community of $j$. The node $i$ is then placed in the community for which this gain is maximum (in case of a tie we use a breaking rule), but only if this gain is positive. If no positive gain is possible, $i$ stays in its original community. This process is applied repeatedly and sequentially for all nodes until no further improvement can be achieved and the first phase is then complete. [...] The second phase of the algorithm consists in building a new network whose nodes are now the communities found during the first phase. To do so, the weights of the links between the new nodes are given by the sum of the weight of the links between nodes in the corresponding two communities. Once this second phase is completed, it is then possible to reapply the first phase of the algorithm to the resulting weighted network and to iterate. Let us denote by "pass" a combination of these two phases. By construction, the number of meta-communities decreases at each pass, and as a consequence most of the computing time is used in the first pass. The passes are iterated [...] until there are no more changes and a maximum of modularity is attained.

Opting for an unweighted network, we used an open-source Python implementation of this algorithm, developed by Thomas Aynaud.[4]

Because the order in which the nodes are evaluated may affect the outcome, we performed the partitioning process for 300 different random orderings, selecting the partition that resulted in the greatest modularity. In our case, a modularity of 0.525158 was achieved.

Some authors contribute to more than one research community. This may happen because the author's research focus is of interest to multiple communities, or because the author has published on several different topics. Regardless of the reason, the employed community detection algorithm will place such authors in the community to which they are most tightly connected. Such authors' will strengthen the relations between the concerned communities.

### 2.7. Community graph

Using the spring layout algorithm in the NetworkX Python library,[5] a graph (Fig. 2) was generated, where nodes correspond to communities, node size to community size, and edge width and node distance depend on inter-community coupling. In most cases, the name of each community was given by the most influential unique keyword, i.e. the top keyword that is used in the articles with the most citations. In a few cases, however, the names were changed by the authors so that the topic of the community could be better reflected, but even in those cases a topic from the rest of the most influential community keyword list was selected. The only exception to this are the names of the cryptography I and II communities that were assigned in an arbitrary manner based on the contained articles. (The following common keywords were not unique to any community: *cyber security, cyber-attacks, security breaches, security, information security, cybersecurity, computer security, cyber threats, network security* and *intrusion detection systems*.)

[4] https://github.com/taynaud

[5] https://networkx.github.io

**Table 1**
Most-cited articles globally.

| Author | Paper title |
|---|---|
| Rivest, R. | A Method for Obtaining Digital Signatures and Public-Key Cryptosystems (1978) [4] |
| Boneh, D. | Identity-based encryption from the weil pairing (2001) [5] |
| Diffie, W. | New Directions in Cryptography (1976) [6] |
| Menezes, A. | Handbook of Applied Cryptography (1996) [7] |
| Shamir, A. | How to Share a Secret (1979) [8] |
| Kocher, P. | Differential power analysis (1999) [9] |
| Shamir, A. | Identity-Based Cryptosystems and Signature Schemes (1985) [10] |
| ElGamal, T. | A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms (1985) [11] |
| Bellare, M. | Random oracles are practical: a paradigm for designing efficient protocols (1993) [12] |
| Paillier, P. | Public-key cryptosystems based on composite degree residuosity classes (1999) [13] |

### 2.8. Sub-community detection and description

For each community, the above process was repeated: An author graph was generated, sub-communities were detected using the Louvain algorithm, sub-community graphs were generated (Figs. 4–26), and each sub-community was summarized in terms of most common keywords, most-cited authors, top publication fora, etc.

## 3. Data

In this section, information regarding the employed raw data is presented.

In total, we have 59,782 articles published from 1949 until early 2020 that are authored by 98,373 authors fully recorded in the database; they contain 148,202 keywords. We also have 835,664 articles recorded as citations (i.e., we only have article title, author surname, and publication year). All the data was acquired from the Scopus database at the end of February 2020. The most-cited articles among them are presented in Table 1, whereas the most common keywords are the following:

1. security
2. privacy
3. authentication
4. cloud computing
5. cryptography
6. cyber security
7. intrusion detection (systems)
8. anomaly detection
9. internet of things (iot)
10. access control

The top five affiliations (in terms of number of papers produced) are: Massachusetts Institute of Technology (including MIT Computer Science and Artificial Intelligence Laboratory), University of California Berkeley, Carnegie Mellon University, Purdue University, and Shanghai Jiao Tong University. Finally, the five most-cited country affiliations are: the United States, China, India, United Kingdom and Germany.

Finally, the top ten publication fora globally are presented in Table 2. It is worth noting that only four out of the top ten publication outlets are conferences while the rest are journals. Therefore, the majority of the collected articles is found on journal publications.

## 4. Results & analysis

In this section, we present the identified community clusters. In total, 12 communities were identified, as shown in Fig. 2. The presentation of each community follows the same structure. First, we address the community topic. In this section, we provide an overview of the most prominent topics in the community. It must be reminded that the clustering is based on individuals and their papers' referencing. Thus we cannot claim that clusters really represent topics in some formal way (as it is done when using topic modeling), every individual can of course cover multiple topics throughout a career, but we do however find a fair cohesiveness with respect to the topics within the communities that is interesting to report. The ambition here is thus to convey an intuitive feel for the topic(s) of the community. To this end, we

**Table 2**
Top publication fora globally.

| Publication forum |
|---|
| Proceedings of the ACM Conference on Computer and Communications Security |
| IEEE Transactions on Information Forensics and Security — Journal |
| Computers and Security — Journal |
| Future Generation Computer Systems — Journal |
| Multimedia Tools and Applications — Journal |
| Proceedings — IEEE Symposium on Security and Privacy |
| Proceedings of SPIE — The International Society for Optical Engineering |
| Information Sciences — Journal |
| Sensors (Switzerland) — Journal |
| Proceedings — IEEE INFOCOM |

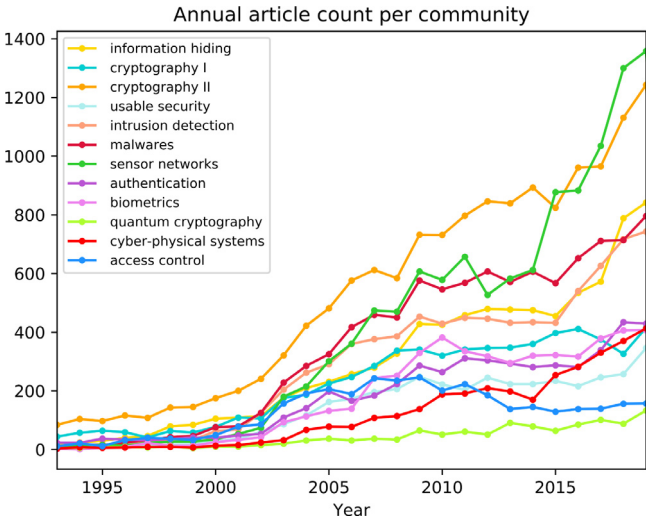


**Fig. 3.** Growth of the detected communities over time.

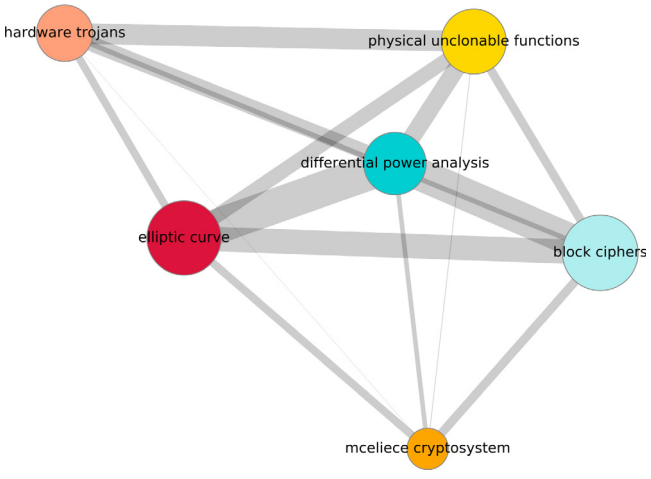


**Fig. 4.** Cryptography I sub-community graph.

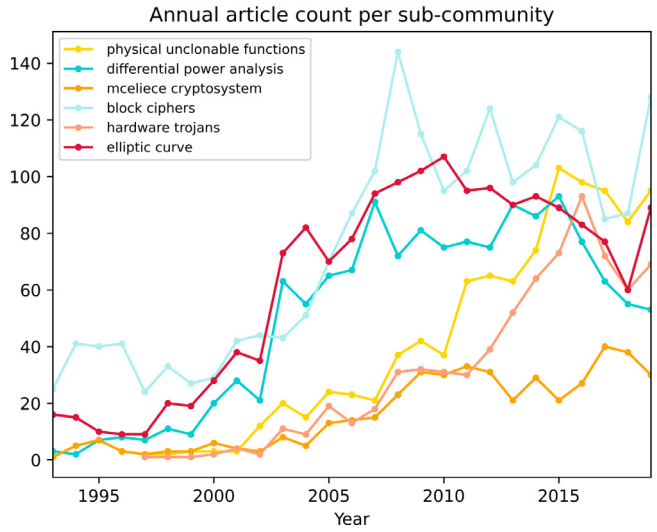


**Fig. 5.** Growth of the detected cryptography I sub-communities over time.

list the most-cited articles produced by members of the community. Furthermore, the sub-communities of each community are also presented and described.

Secondly, properties related to the people and the network of the community are described. Here, we present the most productive countries in the community, the most-cited members of the community, the most popular journals and conferences for dissemination, the most important influences from other communities in terms of most-cited external papers, as well as the historical evolution of the community in terms of papers produced per year. In the dissemination outlet list, general publisher series (e.g., Springer Lecture Notes on Computer Science and ACM International Conference Proceedings Series) have been removed because they represent an excessively large and fuzzy set of conferences and workshops.

Finally, in Fig. 3 the growth of all the detected communities over the last 27 years is presented in terms of the articles included in our analysis. The order in which the communities will be presented below is based on how active each one of them is today, as shown in Fig. 3

### 4.1. Cryptography I & II

The Louvain clustering algorithm identified two communities concerned with cryptography. However, as they are closely related, their joint presentation allows a more coherent description. Even though cryptography has thousands of years of history, the academic discipline emerged in the 1970s with the creation of a public encryption standard (DES) and the invention of public-key cryptography. This community completely dominated cyber security research in the 1980s and 1990s, producing approximately 70% of all published papers in 1985 and 1986. Even though it has maintained its position as the most productive community, and its absolute number of publications continues to rise, its relative share of publications dropped to slightly above 20% in 2018 and 2019. Considering contributing countries, the United States dwarfs all other nations in terms of the number of publications and even more in terms of the number of citations.

At the core of this topic is **provable security**. The corresponding sub-community is concerned with the fundamental mathematical assumptions and abstractions used in cryptography, such as the random oracle model [14] and universal composability [15]. Another sub-community is concerned with **provable data possession**, which is close to **provable security** but focuses more on the fundamentals of data integrity and authenticity verification, and on protocols that provide probabilistic proof that files are stored. As in the case of provable security, the sub-community concerned with **public-key cryptography** has its origins in the 1970s, producing notable contributions such as the RSA cryptographic scheme [4], the ElGamal cryptographic scheme [11], and, later, identity-based encryption [5]. In parallel, a sub-community concerned with symmetric ciphers and cryptanalysis emerged. This sub-community focuses on the construction of **block ciphers**, such as DES [16] and AES [17], as well as on the successful breaking of these cryptographic systems, most notably DES [18].

A common approach to cryptanalysis is to employ side-channel attacks, thus attempting to reveal secrets by measuring unintended side-effects of cryptographic computation. The most interesting side-channel attack is **differential power analysis**, for which there is a dedicated sub-community. Detecting small variations in power consumption patterns during cryptographic operations can be used to find secret keys from otherwise tamper-resistant devices [9]. As proposed by Agrawal et al. [19], side-channel information can be used to detect **hardware trojans**, that is, malicious alterations to integrated circuits [20]. In

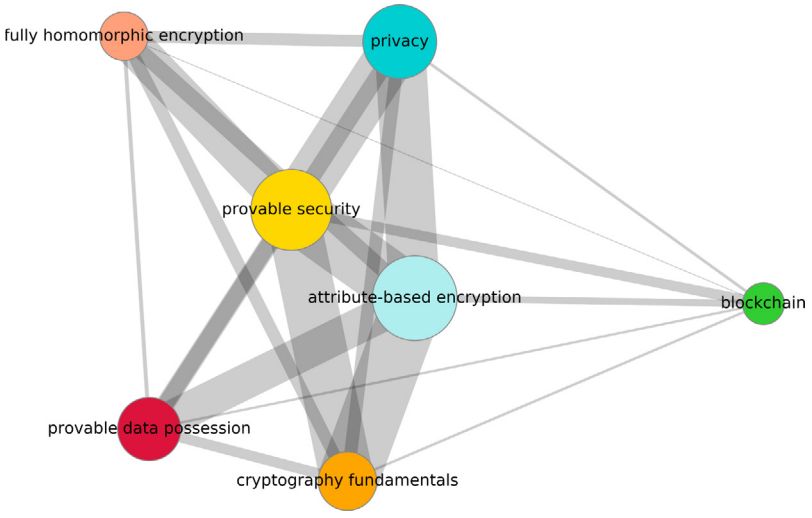


**Fig. 6.** Cryptography II sub-community graph.

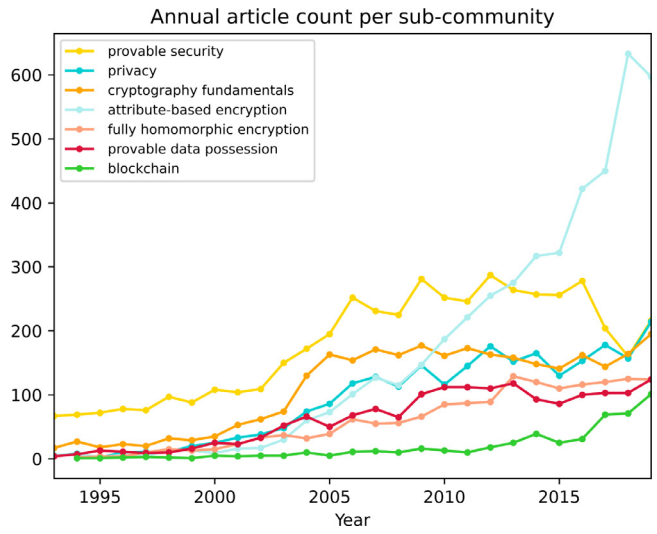


**Fig. 7.** Growth of the detected cryptography II sub-communities over time.

the first decade of the 21st century, a sub-community related to this topic emerged. Other approaches to detecting hardware trojans include the use of **physical unclonable functions** (PUFs). PUFs are primitives for deriving secrets from complex physical characteristics of integrated circuits rather than storing the secrets in digital memory. PUFs make use of random variations during the fabrication process of an integrated circuit, and thus the secret is difficult to predict or extract [21].

The sub-community concerned with **elliptic curve** emerged in the late 1980s, with the independent co-discovery of that cryptographic system by Victor Miller [22] and Neil Koblitz [23].

A useful feature of an encryption system is that it allows operations on the encrypted data without revealing their content. This is the topic of **fully homomorphic encryption** sub-community, dominated by Craig Gentry, the creator of the first fully homomorphic encryption scheme [24]. A sub-community with similarities to the homomorphic encryption group is the one on **privacy preserving** schemes, which is related to issues such as searchable encryption [25] and differential privacy [26].

The sub-community concerned with **attribute-based encryption** has, only the last decade, dominated the cryptographic community. As in the case of homomorphic and privacy-preserving encryption, attribute-based encryption aims to develop methods that allow multiple users to access different parts or aspects of the encrypted data. This is achieved by using attributes to describe the encrypted data or user credentials [27]. The growth of this sub-community has been staggering, amounting to almost 40% of all cryptography publications in 2018 and 2019.

Finally, the most recently appeared sub-community is concerned with **blockchain**, which is directly connected with cryptography, as it is a growing collection of blocks, effectively a chain, each of which is based on the cryptographic hash of the previous block. This sub-community started to make significant contributions around 2012.

### 4.2. Sensor networks

Sensor networks are currently attracting great attention; they represent the largest and most active community (if we consider cryptography I and II separately) in our analysis. This community appeared in the late 1970s (i.e., 1978 and 1979). In 1980, the United States Defense Advanced Research Projects Agency started the Distributed Sensor Network program to explore the challenges in implementing distributed/wireless sensor networks [28].

Since its appearance, the community has followed a continuous increase in publications. In 2002, the community started growing rapidly until 2011, and in 2014 the same growth resumed.

Fig. 8 shows the nine sub-communities. We can further divide them into topics. First, we have applications: **internet of things**, **vehicular communications**, and **implantable devices**. Then, we have network technologies: **wireless sensor networks** and **ad hoc networks**. Finally, we have security mechanisms and attacks: **physical layer security**, **content-based security**, **jamming**, and **spoofing attacks**.

The most active sub-community overall is that concerned with the **internet of things**. The term "internet of things" was introduced around 1999. One of the earliest important articles of this community, however, appeared in 2005 [29] and describes an end-to-end security architecture for constrained embedded devices.

As all the top five community articles (Table 7) are related to security mechanisms, it is no surprise that the sub-community concerned with **physical-layer security** is also on the top of the list of the most productive sub-communities. The most cited article produced by this sub-community is [30], in which the problem of confidentiality over wireless channels is mathematically formulated. Another common problem in sensor networks is how new nodes can be added to the network and be able to

**Table 3**
Most-cited articles produced by the cryptography communities.

| Author | Paper title |
|---|---|
| Boneh, D. | Identity-based encryption from the weil pairing (2001) [5] |
| Bellare, M. | Random oracles are practical: a paradigm for designing efficient protocols (1993) [12] |
| Shamir, A. | Identity-Based Cryptosystems and Signature Schemes (1985) [10] |
| Shamir, A. | How to Share a Secret (1979) [8] |
| Rivest, R. | A Method for Obtaining Digital Signatures and Public-Key Cryptosystems (1978) [4] |

**Table 4**
Top publication fora in the cryptography communities.

| |
|---|
| Proceedings of the ACM Conference on Computer and Communications Security |
| IEEE Access |
| Information Sciences |
| Future Generation Computer Systems |
| IEEE Transactions on Information Forensics and Security |

**Table 5**
Most-cited authors (top 5) in the cryptography communities.

| Author | Citations |
|---|---|
| Boneh, Dan | 8076 |
| Waters, Brent R. | 7414 |
| Shamir, A. | 7368 |
| Sahai, Amit | 6057 |
| Bellare, Mihir | 5592 |

**Table 6**
Most-cited countries (top five) in the cryptography community.

| Country | Citations |
|---|---|
| United States | 271064 |
| China | 44484 |
| Israel | 36752 |
| France | 27303 |
| Germany | 24381 |

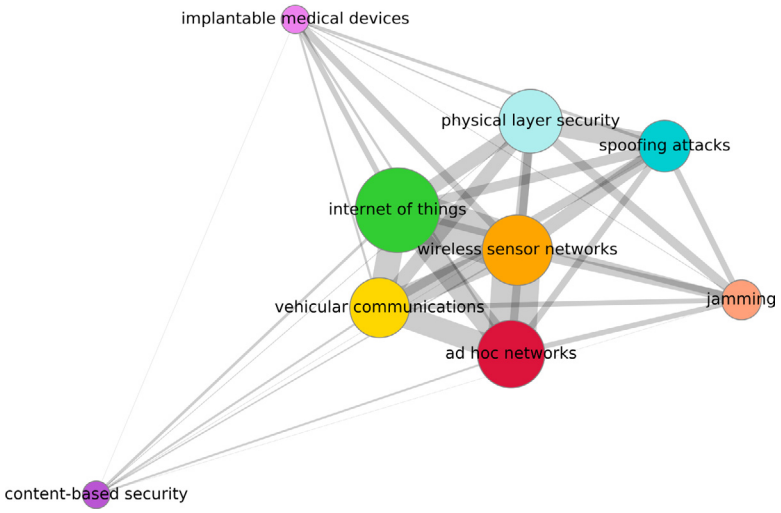


**Fig. 8.** Sensor networks sub-community graph.

**Table 7**
Most-cited articles produced by the sensor networks community.

| Author | Paper title |
|---|---|
| Perrig, A. | SPINS: Security protocols for sensor networks (2002) [33] |
| Eschenauer, L. | A key-management scheme for distributed sensor networks (2002) [31] |
| Karlof, C. | Secure routing in wireless sensor networks: Attacks and countermeasures (2003) [34] |
| Chan, H. | Random key predistribution schemes for sensor networks (2003) [35] |
| Douceur, J. R. | The sybil attack (2002) [36] |

**Table 8**
Top publication fora in the sensor networks community.

| |
|---|
| IEEE Access |
| IEEE Transactions on Vehicular Technology |
| IEEE Internet of Things Journal |
| IEEE Communications Magazine |
| IEEE Transactions on Wireless Communications |

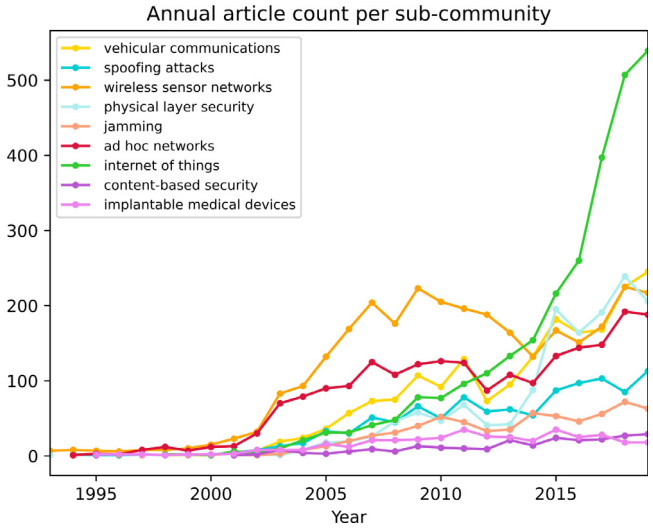


**Fig. 9.** Growth of the detected sensor networks sub-communities over time.

communicate securely with the existing ones. To resolve this, a key-management system is required, as that described in [31]. Then, a **content-based security** mechanism can be used so that each wireless sensor node can only have access to specific content even though the messages are available to all the nodes.

Another topic that is currently attracting attention is **vehicular communications**, where one is interested in the communication between vehicles and between vehicles and the road infrastructure. This area started developing dramatically after 2008 and continues to grow owing to its relation to autonomous vehicles and smart cities. The most cited article produced by this sub-community is [32].

The most influential affiliation country is the United States leading with a big difference from the second which is China while Canada is following closely in the third place. Then Switzerland and United Kingdom also follow with a distance gap.

The **sensor networks** community is closely related to the **cryptography**, **malwares**, and **intrusion detection** communities. This can be explained by the need for authentication and encryption methods in sensor networks. This relation can also be seen from the existence of the **physical-layer security** sub-community within the **sensor networks** community.

**Table 9**
Most-cited authors (top five) in the sensor networks community.

| Author | Citations |
|---|---|
| Perrig, Adrian | 5477 |
| Shen, Xuemin | 2599 |
| Ning, Peng | 1955 |
| Lin, Xiaodong | 1945 |
| Lu, Rongxing | 1816 |

**Table 10**
Most-cited countries (top five) in the sensor networks community.

| Country | Citations |
|---|---|
| United States | 87336 |
| China | 17937 |
| Canada | 13599 |
| Switzerland | 6357 |
| United Kingdom | 6126 |

### 4.3. Information hiding

The **information hiding** community is to a large extent interested in *steganography* and this is the same topic that represents the history and background of this community. Steganography

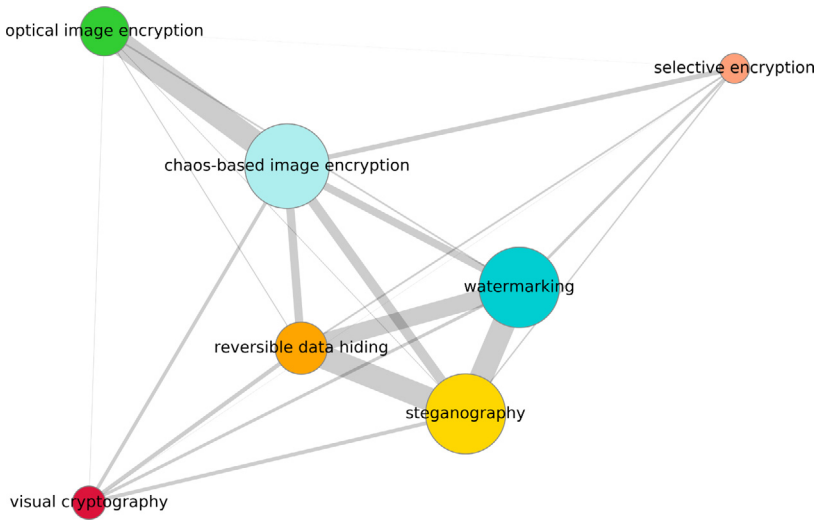


**Fig. 10.** Information hiding sub-community graph.

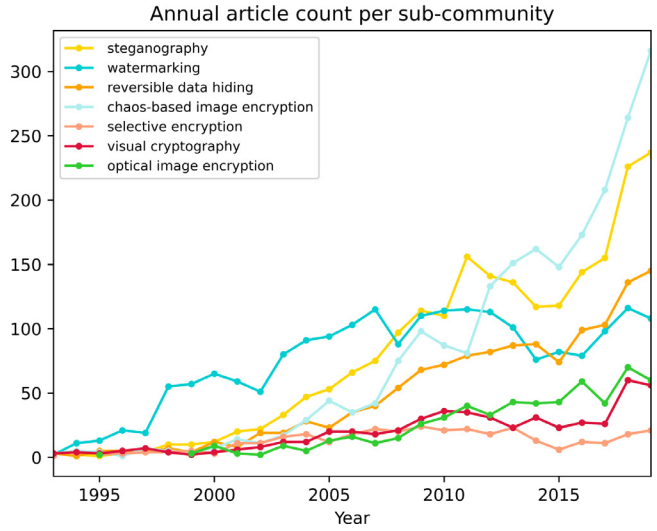


**Fig. 11.** Growth of the detected information hiding sub-communities over time.

is concerned with disguising information in data available to unwanted eavesdroppers. In contrast with cryptography, where it is evident that there is a message sent, in steganography the challenge is to conceal the transmission of a message. This subject stems from information theory. The general principle is to identify redundant bits in data in a *cover medium* and to encode the secret message in a produced *stego medium*.

As in the case of cryptography, the concept of steganography dates back long in history, with examples from ancient Greece, Rome, and China. As a scientific discipline, its foundation was laid in Shannon's paper "Communication Theory of Secrecy Systems" [37], which was published in 1949. To complement the field of cryptography, Shannon introduced "[…] true secrecy systems where the meaning of the message is concealed by cipher, code, etc., although its existence is not hidden, and the enemy is assumed to have any special equipment necessary to intercept and record the transmitted signal".

Despite its early birth, the community had only minor activity until the 1990s. However, since then, it has steadily grown into a large community, with its publication trend pointing upward.

**Steganography** represents the largest sub-community within this community, and it is concerned with both encoding and information hiding. As already mentioned, this is the core of this community and the first, historically, topic of interest. It gained momentum in the late 1990s and has since exhibited a steady increase in article production.

Normally regarded as a complementary approach to steganography, **watermarking** has become one of the largest information hiding sub-communities. The term watermarking relates to a paper-making technique for keeping track of provenance. Watermarking is similar to steganography in that it embeds and hides information in a source data file. However, it also differs significantly. Watermarking has a robustness requirement: it should not be possible to remove (e.g., by image cropping, scaling, and rotation, or through conversion or compression). A watermark is not necessarily hidden, but, as in the case of Kerkhoffs' principle for cryptosystems, it should be difficult to remove even if the algorithm that generated it is known. The general concept of watermarking has a long history, but digital watermarking was born in the early 1990s [38]. Even though it is possible to describe the difference between the watermarking and the steganography sub-communities, their separation in terms of individuals appears less clear. There are people in the steganography group that have produced articles on watermarking.

In the information hiding community, there are also a number of sub-communities concerned with encryption for information hiding purposes.

The first is **chaos-based image encryption**. This sub-community is concerned with encryption techniques based on chaos theory. This approach is based on that chaotic systems are suitable for encryption, as they are sensitive to initial conditions. Authors in this community also note that Shannon, who became a member of this community, already in 1949 [37] (before the development of chaos theory) outlined the fundamental principles for the domain. Despite the old roots of the sub-community, the number of produced papers increased significantly only in the second half of the 2000s. Currently, this sub-domain is one of the two largest, with China dominating the production.

The sub-community of **selective encryption** is concerned with combining compression/decompression with encryption/decryption for multimedia data (video and audio). A fundamental challenge in this field is that encryption and decryption should be performed on large volumes of data and in real time. To balance this trade-off, videos are encrypted only partially and selectively, hence the name. The community grew with wide availability of the internet and the advent of services such as video-on-demand.

**Table 11**
Most-cited articles produced by the information hiding community.

| Author | Paper title |
|---|---|
| Chen, G. | A symmetric image encryption scheme based on 3D chaotic cat maps (2004) [39] |
| Tian, J. | Reversible Data Embedding Using a Difference Expansion (2003) [40] |
| Cox, I. J. | Secure spread spectrum watermarking for multimedia (1997) [41] |
| Petitcolas, F. A. P. | Information hiding - a survey (1999) [42] |
| Shannon, C. E. | Communication theory of secrecy systems (1949) [37] |

**Table 12**
Top publication fora in the information hiding community.

| |
|---|
| Multimedia Tools and Applications |
| Optics Communications |
| IEEE Transactions on Information Forensics and Security |
| Optics and Lasers in Engineering |
| Proceedings of SPIE - The International Society for Optical Engineering |

**Table 13**
Most-cited authors (top five) in the information hiding community.

| Author | Citations |
|---|---|
| Fridrich, Jiri | 3495 |
| Chang, C. C. | 2267 |
| Wang, Xing-yuan | 1762 |
| Anderson, R. J. | 1692 |
| Kilian, Joe | 1399 |

**Table 14**
Most-cited countries (top five) in the information hiding community.

| Country | Citations |
|---|---|
| China | 61045 |
| United States | 38209 |
| Taiwan | 18324 |
| India | 9728 |
| United Kingdom | 4688 |

**Table 15**
Most-cited articles produced by the intrusion detection community.

| Author | Paper title |
|---|---|
| Denning, D. | An Intrusion-Detection Model (1987) [45] |
| Mchugh, J. | Testing Intrusion Detection Systems (2000) [52] |
| Forrest, S. | Sense of self for unix processes (1996) [43] |
| Sheyner, O. | Automated generation and analysis of attack graphs (2002) [53] |
| Lippmann, R. P. | Evaluating intrusion detection systems (2000) [54] |

It has been and remains a small community but with a fairly steady production rate.

In **visual cryptography**, the fundamental principle for hiding data in images is to divide a secret image into different shadow images, called shares. The shares are devised so that if certain subsets are combined, the original secret image is recovered, whereas individual shares or combinations of unqualified shares contain no information. The community is fairly small and steady-sized, with Taiwan and China in the front.

As the name implies, the sub-community of **optical image encryption** is concerned with optical filters that diffuse the original image to noise, and then recover it back. These diffusers operate both in the space as well as in the spatial frequency domains, the latter using various mathematical transforms, such as Fourier, Fresnel, and Gyrator. It should be noted that this sub-community is perhaps best considered to belong to an optics community (not studied here) rather than to computer security.

Finally, there is a sub-community concerned with **reversible data hiding**. This group focuses on techniques that insert information by modifying the original file or signal, but they enable the exact restoration of the original after the extraction of the embedded information. A few articles in the community date back to the 1990s; however, our data suggest that it materialized in the second half of the first decade of the 21st century, and it is now established as a small community, with China in the lead.

In general, China dominates the information hiding community, and Taiwan has also a strong position. The US is the second most influential country. Not surprisingly, it has strong academic relationships with cryptography.

### 4.4. Intrusion detection

This community came into being in the late 1990s. It has since experienced uninterrupted growth in productivity, and it was one of the five most productive communities in 2019. The community initially focused on general intrusion/**anomaly detection** systems and **attack graphs**. Important early articles were [43], which presented a method for anomaly detection, and [44], concerned with modeling the relations between various attacker actions. The sub-community concerned with **machine learning anomaly detection** appeared at the same time and focuses on a topic similar to that of the **anomaly detection** sub-community. This is expressed in Dorothy Denning's paper *An Intrusion Detection Model* from 1987 [45]: "[…] the hypothesis that security violations can be detected by monitoring a system's audit records for abnormal patterns of system usage". The machine learning anomaly detection sub-community is the dominant sub-community as of the writing of this article, producing approximately 30% of the total number of articles in the intrusion detection community.

Around 2000, a sub-community grew around a particular type of network attacks, namely, **distributed denial-of-service attacks**, as exemplified by [46]. In the middle of the first decade of the 21st century, as the interest in attack graphs, anomaly detection, and DDoS attacks increased, new sub-communities also emerged. A sub-community developed around signature-based **deep packet inspection**, as described in, for example, [47]. Another sub-community, which also appeared at the same time, is concerned with **information visualization** and developed methods for visualizing security-related network data to facilitate manual intrusion detection, as exemplified by [48].

Around 2010, the interest in **cloud computing** reached its peak. This is one of only two exceptions to the American dominance of the intrusion detection community, as the most influential country, in terms of citations, in the **cloud computing** sub-community is India. A characteristic article is [49], in which different intrusion techniques affecting availability, confidentiality, and integrity of cloud resources and services are surveyed. The second sub-community in which the US is not dominant, which also peaked around 2010, is concerned with **traffic classification** using machine learning. Its focus appears to be the same as that of the **anomaly detection** sub-community. Here, Canada and Spain are among the most influential countries, and publication forums are generally concerned more with topics related to networks and communications.

Since 2010, the sub-community concerned with **software-defined networking** (SDN) has significantly increased in terms output. It focuses on multiple security concerns in the SDN domain, including intrusion detection (e.g., [50]) as well as control-plane saturation attacks [51].

Most closely connected with the **malwares** community, the **intrusion detection** community has also connections with the **sensor networks** community.

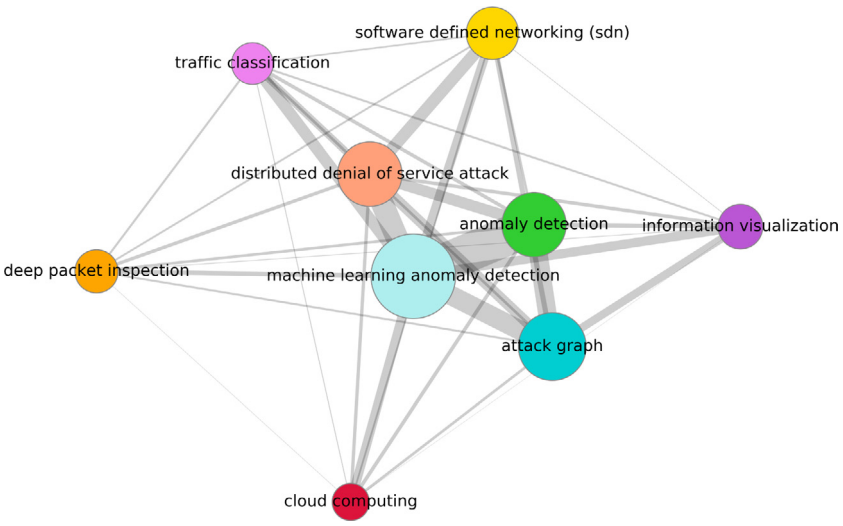


**Fig. 12.** Intrusion detection sub-community graph.

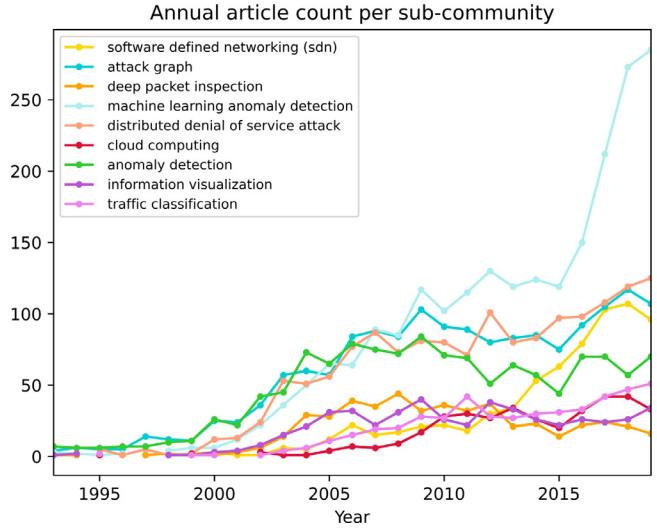


**Fig. 13.** Growth of the detected intrusion detection sub-communities over time.

**Table 16**
Top publication fora in the intrusion detection community.

| |
|---|
| Computers and Security |
| IEEE Access |
| Computer Networks |
| Journal of Network and Computer Applications |
| Expert Systems with Applications |

**Table 17**
Most-cited authors (top five) in the intrusion detection community.

| Author | Citations |
|---|---|
| Jajodia, Sushil | 2091 |
| Stolfo, Salvatore J. | 1124 |
| Forrest, Stephanie | 935 |
| Denning, Dorothy E. | 778 |
| Lippmann, R. P. | 658 |

### 4.5. Malwares

The **malwares** community has been active since 1973. Malware research is focused on discovering, preventing, and stopping malicious software, including viruses, trojans, ransomware, and spyware. Early papers produced by members of this community were related to secure information flow [55] and the modeling of security policies [56]. Thus, this community is closely related to other cyber security communities such as intrusion detection. More fundamental work was carried out slightly later, with, for example, Fred Cohen from Lehigh University, presenting early theory and experiments on computer viruses [57]. Another influential paper (from 1991) used directed-graph epidemiological models for the spread of computer viruses [58].

**Table 18**
Most-cited countries (top five) in the intrusion detection community.

| Country | Citations |
|---|---|
| United States | 58718 |
| China | 7359 |
| Australia | 5762 |
| India | 4579 |
| Canada | 4196 |

The community started publishing slowly, with a few papers per year in the 70s and 80s. This increased up to a few hundred of papers per year in the 90s and continued to rise in the 2000s until peaking with 695 papers in 2018. Although there are some early influential papers, the most cited were published around 2010, with the most cited paper being "TaintDroid: An information-flow tracking system for real-time privacy monitoring on smartphones" (2014) by William Enck et al. [59].

Fig. 14 shows the eight **malware** sub-communities. The two largest are **malware detection** and **android**, which are currently the most active. The most cited paper in **malware detection** is [60] and represents the community well, with its focus on finding patterns and detecting new malicious executables. Regarding the **android** community, the focus is on finding malware in Android applications, which are open source and thus suitable for analysis. In the most cited paper of this sub-community, the authors collected more than 1200 malware samples in the Android platform and tested four security applications, demonstrating that these could only detect 20%–80% of the existing malware [61]. Both sub-communities are dominated by researchers active in the US, followed by Germany.

One sub-community is concerned with the topic of malicious network **traffic analysis**, as for example communication between botnets and malwares. The **information flow** sub-community is concerned with information flow analysis, which is the information theoretical study of securing data used by computer systems.

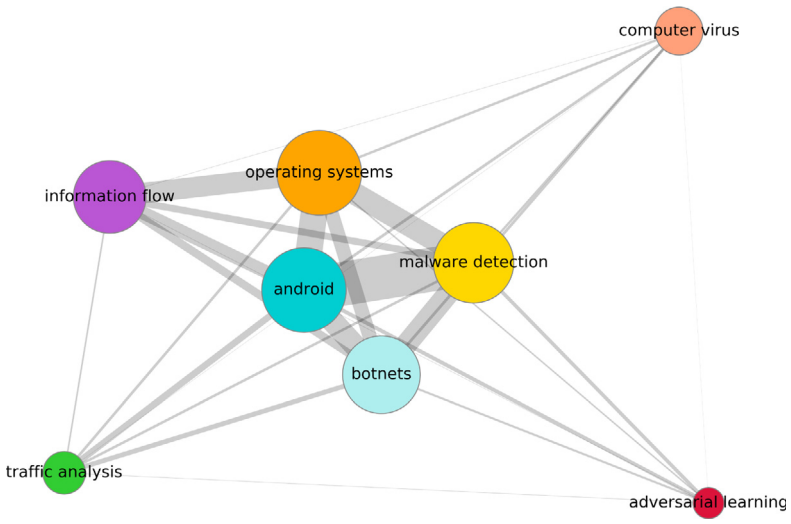


**Fig. 14.** Malwares sub-community graph.

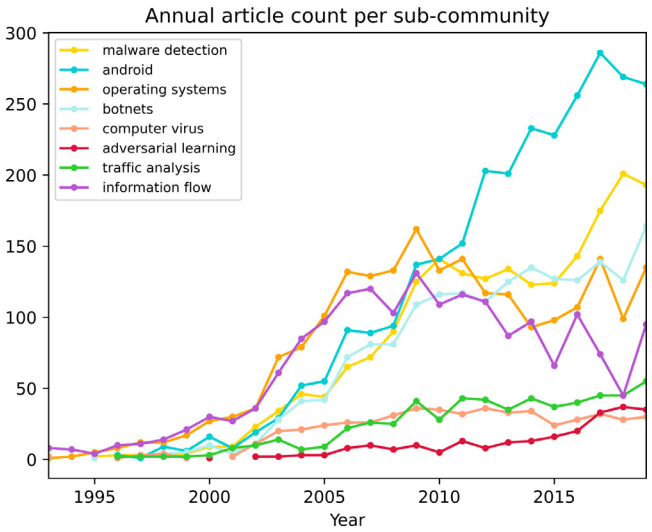


**Fig. 15.** Growth of the detected malwares sub-communities over time.

**Table 19**
Most-cited articles produced by the malwares community.

| Author | Paper title |
|---|---|
| Enck, W. | TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones (2014) [59] |
| Zhou, Y. | Dissecting Android malware: Characterization and evolution (2012) [61] |
| Sabelfeld, A. | Language-based information-flow security (2003) [63] |
| Felt, A. P. | Android permissions demystified (2011) [64] |
| Enck, W. | On lightweight mobile phone application certification (2009) [65] |

The study of **botnets**, **computer viruses**, and virtualization techniques for **operating systems** that enhance security, follow with slightly less publications. Finally, **adversarial learning** represents a novel research field that came into existence only after 2000 and employs machine learning techniques to model adversaries so that attack simulations can then be performed, as described in [62].

The **malwares** community is closely related to the **intrusion detection**, **sensor networks**, and **cryptography** communities.

**Table 20**
Top publication fora in the malwares community.

| |
|---|
| Proceedings of the ACM Conference on Computer and Communications Security |
| Proceedings - IEEE Symposium on Security and Privacy |
| Computers and Security |
| Proceedings - Annual Computer Security Applications Conference, ACSAC |
| IEEE Access |

**Table 21**
Most-cited authors (top five) in the malwares community.

| Author | Citations |
|---|---|
| Wagner, David | 3767 |
| Song, Dawn Xiaodong | 3562 |
| Kruegel, Christopher | 2090 |
| Lee, Wenke | 2059 |
| McDaniel, Patrick | 1511 |

**Table 22**
Most-cited countries (top five) in the malwares community.

| Country | Citations |
|---|---|
| United States | 124853 |
| Germany | 14456 |
| China | 9242 |
| Italy | 8035 |
| United Kingdom | 4871 |

### 4.6. Biometrics

The **biometrics** community is one of the largest communities in our analysis, in terms of community members. It appeared in the early 1980s, almost two decades after the introduction of the first semi-automatic face recognition system by Woodrow Bledsoe in 1968.

It followed a slow but steady productivity growth; currently, it is exactly in the middle among all communities in terms of productivity.

As seen in Fig. 16, the **biometrics** community has eight sub-communities, which can be divided into three research domains. The first focuses on applications: **biometric fingerprinting** and **surveillance systems**. The second is concerned with authentication schemes: **keystroke dynamics** and **biometric authentication**. The third is concerned with methods: **face recognition**, **gait recognition**, **person re-identification**, and **background subtraction**.

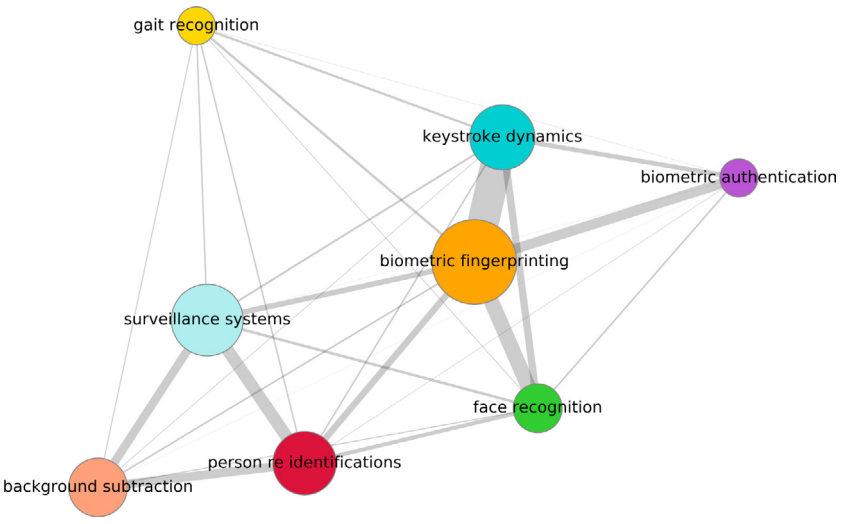


**Fig. 16.** Biometrics sub-community graph.

Overall, the most active and oldest sub-community is that concerned with **biometric fingerprinting**. Articles in this community primarily focus on the general design of biometric systems and their procedures, as for example in [66]. The most interesting older article is [67], which is a study on secure, off-line, authenticated user-identification schemes based on a biometric system.

The **surveillance systems** sub-community is not only related to identifying persons but also to privacy concerns regarding such systems, as presented in [68].

The **biometric fingerprinting** sub-community is closely related to both the **biometric authentication** and the **keystroke dynamics** sub-community. The **biometric authentication** sub-community is concerned with all possible types of biometric features, such as neural activity and brainwaves. Interestingly, the **keystroke dynamics** sub-community began publishing in 1990 and is currently the second most active sub-community. One of the earliest and most cited articles is [69], which described a user authentication/identification method by studying keyboard typing habits.

Another research topic in biometrics that is currently attracting great attention is **person re-identification**, which is the process of associating images of a person captured from different cameras or from the same camera in different environments. As expected, this is also related to **face recognition**.

Finally, **background subtraction** is a technique that removes the background of an image or video to study only useful content, something that is used in biometrics recognition. **Gait recognition** is the study of human motion, which can be considered a biometric feature, and can be used to identify people.

The most influential affiliation country is once more, the United States leading with a significant difference from the second one which is China, while Italy is following closely. Then United Kingdom and South Korea follow in some distance.

The **biometrics** community appears distantly related to the other communities in our analysis, but it is closer to the **information hiding** community.

### 4.7. Cyber–physical systems

The **cyber–physical systems** community is a medium-sized community, which is relatively new, as it came into existence approximately 25 years ago. It has experienced a steady growth and is currently the seventh most productive community.

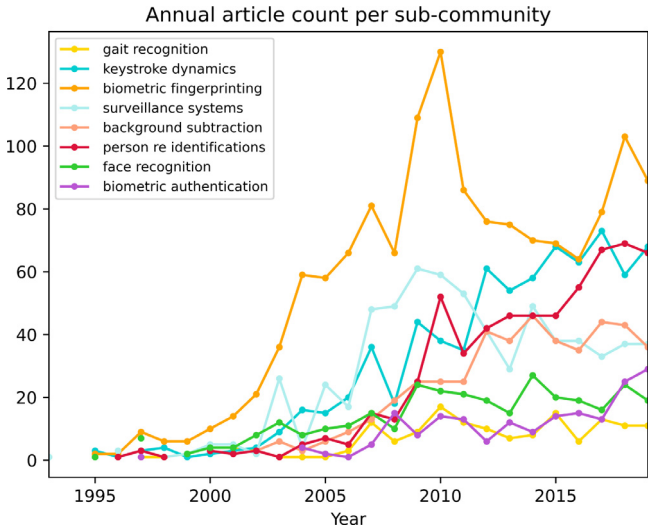


**Fig. 17.** Growth of the detected biometrics sub-communities over time.

**Table 23**
Most-cited articles produced by the biometrics community.

| Author | Paper title |
|---|---|
| Jain, A. K. | An Introduction to Biometric Recognition (2004) [70] |
| Juels, A. | Fuzzy commitment scheme (1999) [71] |
| Ratha, N. K. | Enhancing security and privacy in biometrics-based authentication systems (2001) [72] |
| Ratha, N. K. | Generating cancelable fingerprint templates (2007) [73] |
| Uludag, U. | Biometric cryptosystems: Issues and challenges (2004) [66] |

**Table 24**
Top publication fora in the biometrics community.

| |
|---|
| IEEE Transactions on Information Forensics and Security |
| IEEE Transactions on Circuits and Systems for Video Technology |
| Pattern Recognition |
| Proceedings of SPIE - The International Society for Optical Engineering |
| Pattern Recognition Letters |

Fig. 18 indicates two main research domains. The first is concerned with existing infrastructures and possible methods for

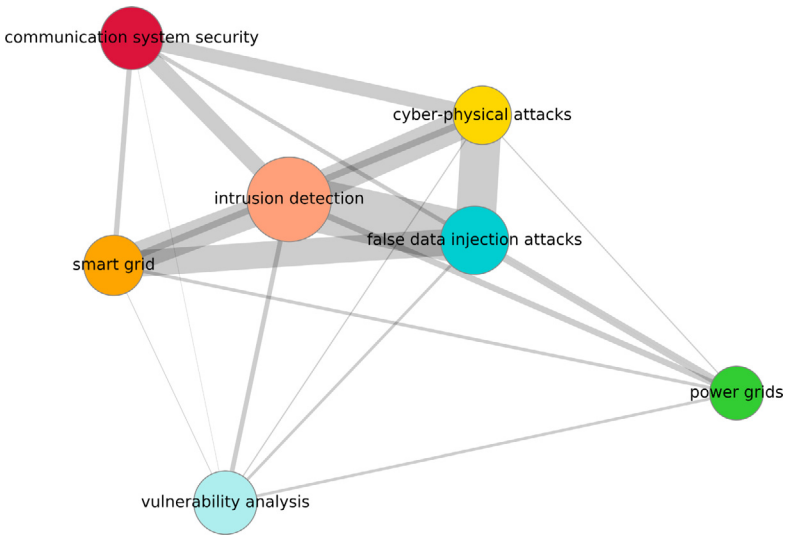


**Fig. 18.** Cyber–physical systems sub-community graph.

**Table 25**
Most-cited authors (top five) in the biometrics community.

| Author | Citations |
|---|---|
| Jain, Anil K. | 1342 |
| Pankanti, Sharath | 624 |
| Bolle, Ruud | 524 |
| Ross, Arun | 501 |
| Connell, J. H. | 490 |

**Table 26**
Most-cited countries (top five) in the biometrics community.

| Country | Citations |
|---|---|
| United States | 20721 |
| China | 5364 |
| Italy | 4194 |
| United Kingdom | 2411 |
| South Korea | 1912 |

**Table 27**
Most-cited articles produced by the cyber–physical systems community.

| Author | Paper title |
|---|---|
| Liu, Y. | False data injection attacks against state estimation in electric power grids (2009) [75] |
| Koscher, K. | Experimental security analysis of a modern automobile (2010) [77] |
| Kosut, O. | Malicious data attacks on the smart grid (2011) [78] |
| Sridhar, S. | Cyber-physical system security for the electric power grid (2012) [74] |
| Pasqualetti, F. | Attack detection and identification in cyber–physical systems (2013) [76] |

defending them: **smart grids**, **power grids**, **communication-system security**, and **intrusion detection systems**. The second is concerned with attacks on such systems: **false data injection attacks**, **cyber–physical attacks**, and **vulnerability analysis**.

Owing to the move from simple control systems towards IT systems, the **intrusion detection systems** and **vulnerability analysis** sub-communities, which are primarily IT-related, are found within the **cyber–physical systems** community. The most important article published by the **intrusion detection systems** sub-community, which is the most active one, is [74], in which the significance of cyber infrastructure security within the power domain, to prevent, mitigate, and tolerate cyber-attacks, is highlighted.

The **false data injection attacks** sub-community is currently the second most active. The majority of publications in this sub-community appeared after 2009, and one of the earliest important articles is [75], in which *false data injection attacks* against power grids were introduced. Such attacks are performed when an attacker maliciously introduces crafted errors into certain system state variables with the aim of manipulating the system.

Another of the most active sub-communities is concerned with **cyber–physical attacks**. In the most cited article [76] a mathematical framework for cyber–physical systems, attacks, and monitors is proposed, and then centralized or distributed attack detection and identification systems are designed.

The other infrastructure-related sub-communities are concerned with **smart grids**, **power grids** and **communication-system security**. The term *smart grid* was first defined by the Energy Independence and Security Act of 2007 (EISA-2007) in the US, and around 2010 the first related papers appeared. The **power grids** sub-community adds one more attack vector to the domain, namely, cascading failures, in which the failure of one component leads to the failure of other components in an interconnected system, are primarily studied by this sub-community.

Finally, the most important article in the **communication system security** sub-community is [77]. This article is important because it is an experimental security analysis of a mix of industrial-grade networks and cyber–physical systems, which are found not only in vehicles but also in the energy domain.

The most influential affiliation country is the United States leading with a big difference from the second which is China while the United Kingdom and Sweden are following in the third and fourth position.

One observation is that this community is equally concerned with attacks on state estimators for power grids, as also with attacks on the wider industrial control systems (ICS).

The **cyber–physical systems** community is closely related to the **sensor networks** community, as sensors and sensor networks are becoming a standard in modern power grids. Additionally, it is also related to the **intrusion detection** community. This is due to the fact that **vulnerability analysis** of power grid infrastructures is becoming increasingly widespread.

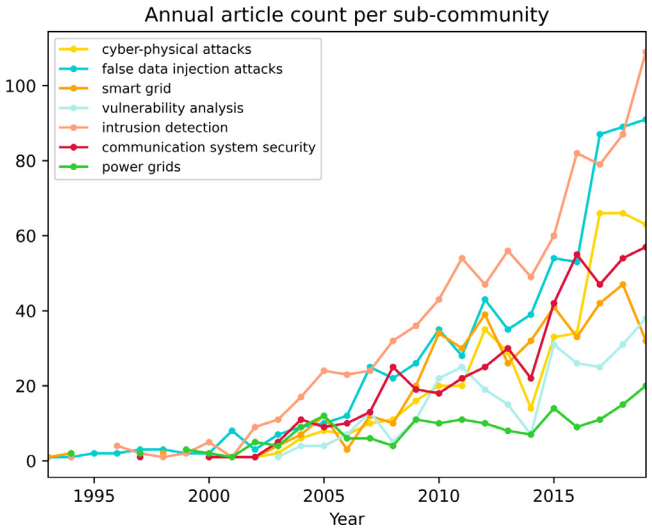


**Fig. 19.** Growth of the detected Cyber-physical systems sub-communities over time.

**Table 28**
Top publication fora in the cyber–physical systems community.

| |
|---|
| IEEE Transactions on Smart Grid |
| IEEE Transactions on Power Systems |
| IEEE Transactions on Industrial Informatics |
| Reliability Engineering and System Safety |
| IEEE Access |

**Table 29**
Most-cited authors (top five) in the cyber–physical systems community.

| Author | Citations |
|---|---|
| Reiter, Michael K. | 1221 |
| Kohno, Tadayoshi | 1002 |
| Liu, Yao | 488 |
| Sastry, S. | 451 |
| Cárdenas, Alvaro A. | 395 |

**Table 30**
Most-cited countries (top five) in the cyber–physical systems community.

| Country | Citations |
|---|---|
| United States | 24938 |
| China | 2886 |
| United Kingdom | 1633 |
| Sweden | 1410 |
| Italy | 1131 |

### 4.8. Authentication

The **authentication** community is a relatively small-size community although it is one of the oldest communities in our analysis. It started in the late 1970s with research on authentication (using passwords) and authenticated encryption systems for computers. The Diffie–Hellman key exchange [6] is a characteristic example.

This community followed a steady growth in productivity, except for the period 2012–2016, during which it remained static. Currently, it is the eighth out of the twelve communities in terms of productivity.

As seen in Fig. 20, it has six sub-communities. Among them, the **mutual authentication** sub-community is currently the most active. It is primarily concerned with "two-factor authentication" (also called mutual authentication), which is commonly achieved by using a hardware authentication device (such as a OTP (one-time password) generator or OTP device). The majority of publications in this sub-community were made after 2006, and one of the earliest important articles was [79], in which a two-factor authentication protocol for wireless sensor networks was proposed.

However, one of the most active sub-communities in the past was the **password** sub-community. A characteristic example is [80], which proposes a secure password authentication method that is immune to eavesdropping and tampering by an attacker. This method, which is currently widely used, involves the use of hashed passwords. In more recent articles, a close relation to the **mutual authentication** sister sub-community can be seen.

**Authentication** mechanisms can also be used in tandem with **confidentiality** mechanisms and achieve **key agreement**; these are two homonymous sub-communities. The subcommunity concerned with **confidentiality** is currently the second most active. Finally, **rfid** (radio-frequency identification) is another hardware solution that can be used as a two-factor authentication token, hence it exists as a sub-community on the authentication community.

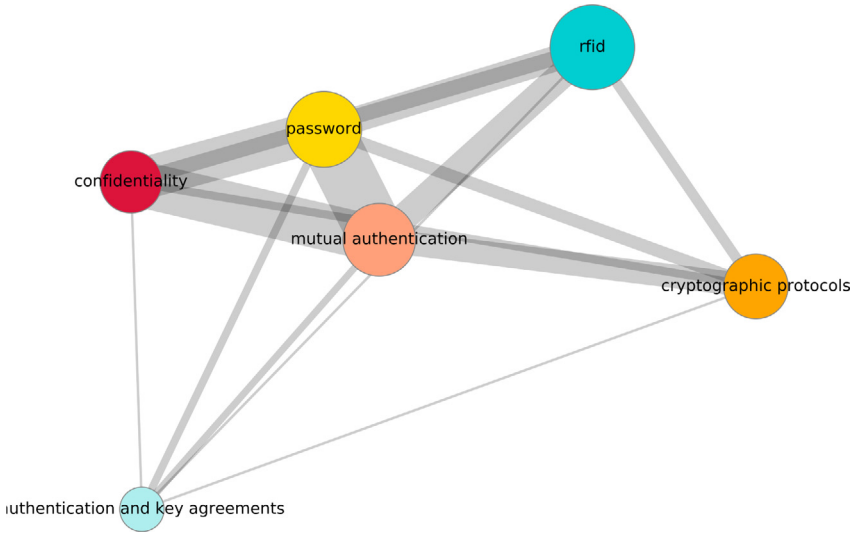


**Fig. 20.** Authentication sub-community graph.

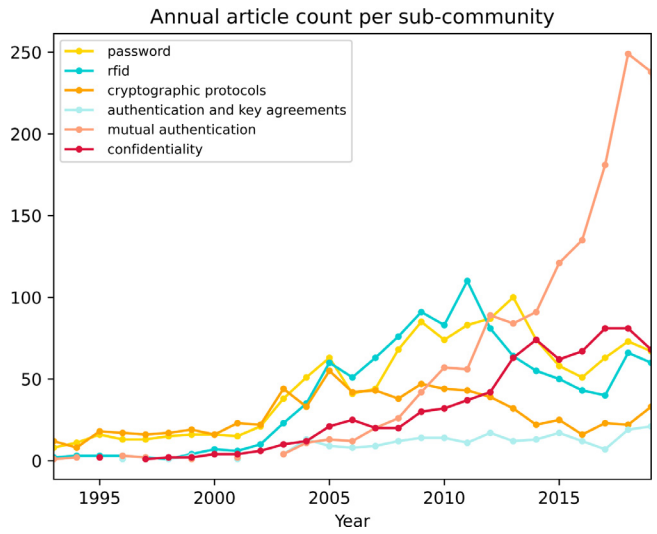


**Fig. 21.** Growth of the detected authentication sub-communities over time.

**Table 31**
Most-cited articles produced by the authentication community.

| Author | Paper title |
|---|---|
| Dolev, D. | On The Security of Public Key Protocols (1981) [81] |
| Diffie, W. | New Directions in Cryptography (1976) [6] |
| Lamport, L. | Password Authentication with Insecure Communication (1981) [80] |
| Messerges, T. S. | Examining smart-card security under the threat of power analysis attacks (2002) [82] |
| Burrows, M. | A logic of Authentication (1990) [83] |

**Table 32**
Top publication fora in the authentication community.

| |
|---|
| Journal of Medical Systems |
| Security and Communication Networks |
| IEEE Access |
| Wireless Personal Communications |
| International Journal of Communication Systems |

**Table 33**
Most-cited authors (top five) in the authentication community.

| Author | Citations |
|---|---|
| Hwang, Min-Shiang | 1906 |
| Diffie, Whitfield | 1786 |
| Abadi, Martin | 1729 |
| Hellman, Martin E. | 1657 |
| Khan, Muhammad Khurram | 1423 |

The most influential affiliation country is China, leading with a small difference from the second one which is Taiwan, while the United States is following closely.

The **authentication** community is closely related to the **cryptography** community. This is because authentication uses cryptographical elements, hence the **cryptographic protocols** sub-community. For example, the Diffie–Hellman key exchange, mentioned previously, uses public-key cryptography for both encryption and authentication. **Authentication** is also closely related to the **sensor networks** community, as sensors and sensors networks require authentication and security methods. This relation can also be seen from the **physical layer security** sub-community within the **sensor networks** community.

**Table 34**
Most-cited countries (top five) in the authentication community.

| Country | Citations |
|---|---|
| China | 21642 |
| Taiwan | 20289 |
| United States | 16858 |
| India | 10270 |
| South Korea | 5889 |

### 4.9. Usable security

The **Usable security** community has been active since 1973, but at that time, it was concerned more with **protection motivation theory**, which aims to clarify fear appeals and proposes that people protect themselves based on a number of different factors. Then, in the late 1980s, the term **phishing** came into existence and research that is more related to **phishing**, **usable security**, and information security awareness began to appear. **Phishing** relates to fraudulent techniques for obtaining sensitive information by disguising as a trustworthy entity. The earliest important article is [84], in which the **protection motivation theory** was founded.

Initially, the community produced a few papers per year, but after 2002, it experienced greater growth. Currently, it is one of the smallest communities in terms of size, and tenth out of twelve in terms of productivity.

Fig. 22 shows the six sub-communities. The two largest ones are concerned with **password security** and **phishing**. The former focuses on the study of password habits, graphical passwords, and other password-related topics.

The **phishing** sub-community is concerned with both phishing and mitigation techniques for **phishing** (anti-phishing), which is a very modern topic of research.

The **protection motivation theory** sub-community, which was historically the largest one until 2009, focuses on information security awareness and information security policy compliance.

The **cybercrime** sub-community appeared in 2002 and is concerned with studying the social networks of malware writers and hackers, the social behavior in online black markets, and the creation of attacker profiles among others. The reason for having such a sub-community is no other than the fact that cybercrime can also be the result of low information security awareness and phishing attacks.

Finally, the **economics** sub-community is concerned with the economic effect of phishing attacks and the economics of security investments, whereas the **trust** sub-community with trust issues in IT systems.

Once more, the most influential affiliation country is the United States, while United Kingdom comes second, and Canada is in the third place. Then Germany and Finland are also following.

The **usable security** community is closely related to the **malwares** and **intrusion detection** communities.

### 4.10. Access control

The **access control** community is currently one of the smallest (in terms of size) and least active. It began publishing in the middle 1970s and was primarily concerned with **role-based access control** and **access-control policies**. One of the most important early articles is [90], which is also the most cited in the community. This article focuses on a certain type of access control, namely, role-based access control (RBAC), and describes a framework in which the use and management of RBAC can become easier and more effective.

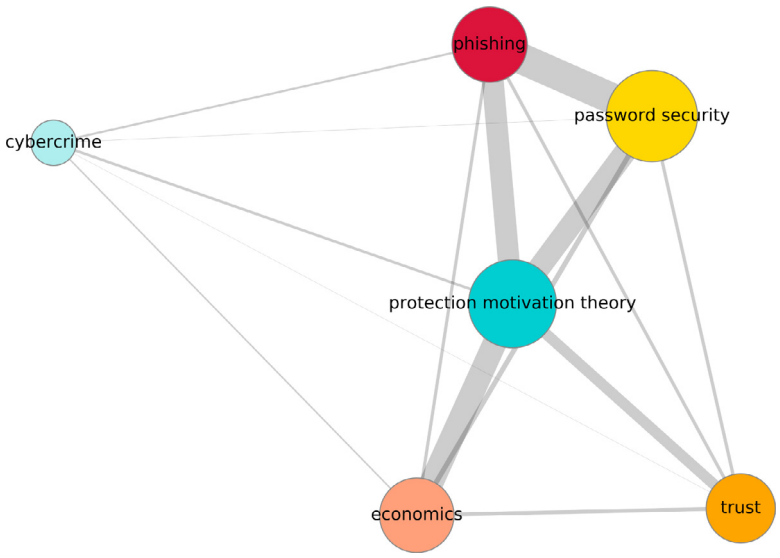


**Fig. 22.** Usable security sub-community graph.

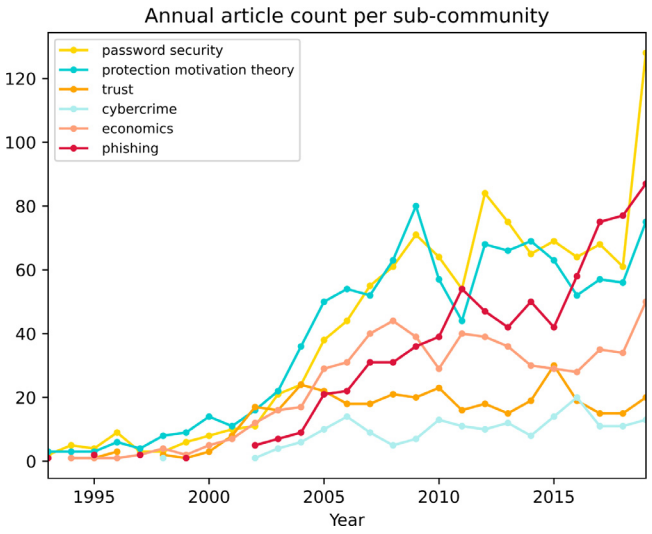


**Fig. 23.** Growth of the detected usable security sub-communities over time.

**Table 35**
Most-cited articles produced by the usable security community.

| Author | Paper title |
|---|---|
| Straub, D. W. | Coping with systems risk: Security planning models for management decision making (1998) [85] |
| Dhamija, R. | Why phishing works (2006) [86] |
| Herath, T. | Protection motivation and deterrence: A framework for security policy compliance in organizations (2009) [87] |
| Bulgurcu, B. | Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness (2010) [88] |
| Johnston, A. C. | Fear appeals and information security behaviors: An empirical study (2010) [89] |

**Table 36**
Top publication fora in the usable security community.

| |
|---|
| Computers and Security |
| Conference on Human Factors in Computing Systems — Proceedings |
| Computers in Human Behavior |
| Proceedings of the ACM Conference on Computer and Communications Security |
| Decision Support Systems |

**Table 37**
Most-cited authors (top five) in the usable security community.

| Author | Citations |
|---|---|
| van Oorschot, Paul C. | 2314 |
| Cranor, Lorrie Faith | 1081 |
| Siponen, Mikko T. | 647 |
| Hong, Jason I. | 574 |
| Furnell, S. M. | 488 |

**Table 38**
Most-cited countries (top five) in the usable security community.

| Country | Citations |
|---|---|
| United States | 35537 |
| United Kingdom | 5444 |
| Canada | 5333 |
| Germany | 2714 |
| Finland | 1503 |

Until 2009, its member count was slowly increasing. Subsequently, it shrank, and in the last six years, it has remained steady.

Fig. 24 shows the six sub-communities. Among them, the **privacy** community has been one of the most active. However, its size has also shrunk, following the parent community. This sub-community is concerned more with trust and privacy issues in software applications, but it also conducts research on policy and privacy management as well as solution enforcement. The most important article is [91], which presented a new, at that time, trust management system, called Policy Maker.

The **security requirements** sub-community is currently the second most active. It is concerned with the study, analysis, and/or modeling of the security and privacy requirements of existing applications, as presented, for example, in [92]. Then, the **context-aware computing** sub-community is concerned with access control mechanisms for ubiquitous computing. Finally, the **grid computing** sub-community is concerned with access-control systems in grid computing.

The most influential affiliation country of the whole community is the United States leading with a significant difference from the second one which is Italy, while United Kingdom is following very closely.

The **access control** community is closely related to the **cryptography** and **malwares** communities. The first relation could be

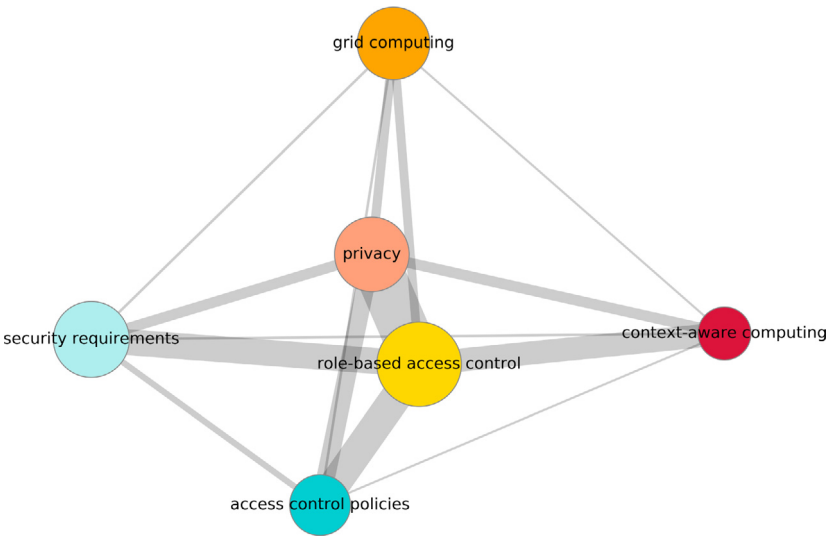


**Fig. 24.** Access control sub-community graph.

**Table 39**
Most-cited articles produced by the access control community.

| Author | Paper title |
|---|---|
| Sandhu, R. | Role-based access control models (1996) [90] |
| Ferraiolo, D. F. | Proposed NIST Standard for Role-Based Access Control (2001) [93] |
| Blaze, M. | Decentralized trust management (1996) [91] |
| Sindre, G. | Eliciting security requirements by misuse cases (2000) [92] |
| Bertino, E. | TRBAC: A Temporal Role-Based Access Control Model (2001) [94] |

**Table 40**
Top publication fora in the access control community.

| |
|---|
| Proceedings of ACM Symposium on Access Control Models and Technologies, SACMAT |
| ACM Transactions on Information and System Security |
| Proceedings of the ACM Conference on Computer and Communications Security |
| Computers and Security |
| Future Generation Computer Systems |

**Table 41**
Most-cited authors (top five) in the access control community.

| Author | Citations |
|---|---|
| Sandhu, Ravinderpal S. | 2054 |
| Bertino, Elisa | 1711 |
| Samarati, Pierangela | 1128 |
| Li, Ninghui | 892 |
| Ahn, Gail-Joon | 738 |

**Table 42**
Most-cited countries (top five) in the access control community.

| Country | Citations |
|---|---|
| United States | 24210 |
| Italy | 5731 |
| United Kingdom | 4174 |
| Germany | 2032 |
| Canada | 1015 |

explained because together with access control an authentication mechanism is needed. Malwares on the other hand are related to access control systems because many times they can bypass them.

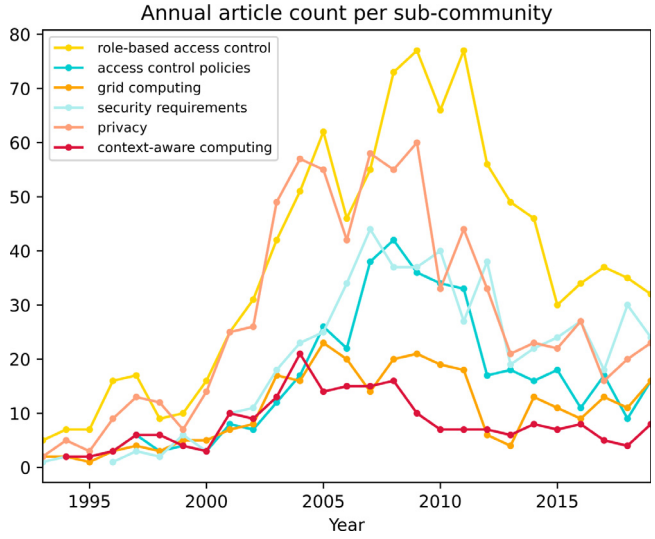


**Fig. 25.** Growth of the detected access control sub-communities over time.

### 4.11. Quantum cryptography

**Quantum cryptography** uses quantum mechanics to perform cryptographic tasks. The best-known example of quantum cryptography is **quantum key distribution**. In our analysis, it corresponds to the smallest and least active community. The community came into existence in the early 1980s. One of the earliest important articles is [95], which is also one of the most cited in the community. In this article, the fundamental requirements for achieving **quantum key distribution** are described. The community has followed a slow but steady growth in productivity.

Fig. 26 shows the three sub-communities. The **quantum secure direct communication** sub-community is the most active. The majority of publications in this sub-community were made after 2004 and are related to achieving secure direct communication, which is secret information that can be transmitted directly through a quantum channel without the use of a private key, as for example in [96].

The **quantum key distribution** sub-community is the second most active and is concerned with key generation and distribution between two parties over quantum communication channels.

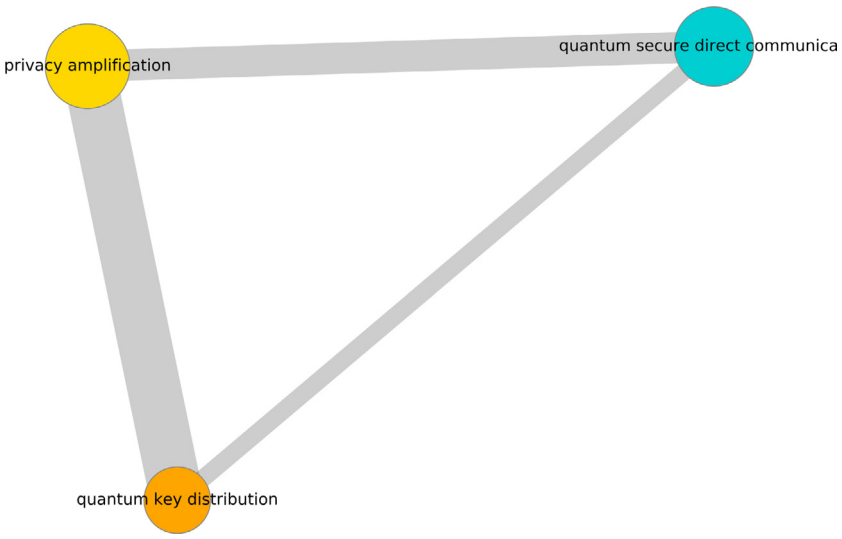


**Fig. 26.** Quantum cryptography sub-community graph.

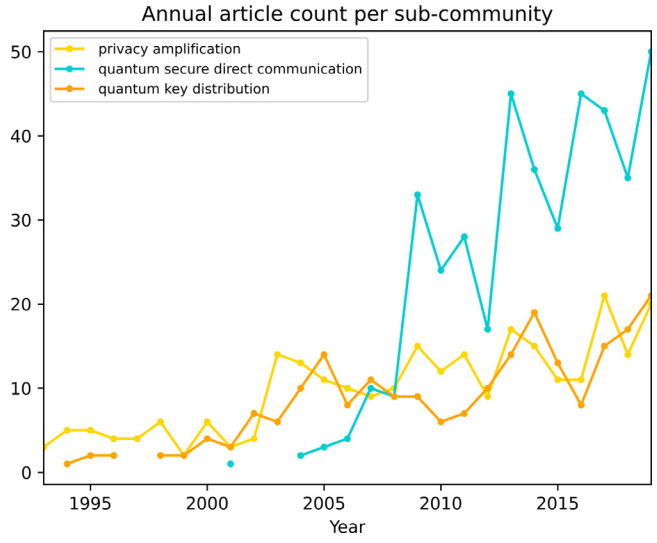


**Fig. 27.** Growth of the detected quantum cryptography sub-communities over time.

The **privacy amplification** sub-community is concerned with encryption techniques using quantum mechanics, as for example in [97].

The absence of a "post-quantum cryptography" sub-community might be obvious but there is an explanation for that. Since homomorphic encryption is usually based on lattice-based methods and post-quantum encryption, the post-quantum sub-community is absorbed and split among the **fully homomorphic encryption** and **McEliece cryptosystem** sub-communities, which are found within the two **cryptography** communities.

The most influential affiliation country is China, second is the United States while Canada is following closely in the third place.

This community is closely related to the **cryptography** communities and less closely related to the **sensor networks** and **steganography** communities.

## 5. Related work

The Cyber Security Body of Knowledge (CyBOK) [101] is an ambitious attempt to identify the foundational knowledge areas of the cyber security sector and inform both academia and practitioners about them. CyBOK differs from the current work in both subject and method. CyBOK aims to organize the cyber security knowledge rather than to understand the research community. It employs consultation workshops with experts and online surveys as compared to the present work's quantitative analysis based on abstract and citation databases. Finally, CyBOK aims for a balance

**Table 43**
Most-cited articles produced by the quantum cryptography community.

| Author | Paper title |
|---|---|
| Bennett, C. H. | Quantum cryptography: Public key distribution and coin tossing (2014) [98] |
| Gisin, N. | Quantum cryptography (2002) [99] |
| Bennett, C. H. | Quantum cryptography using any two nonorthogonal states (1992) [95] |
| Shor, P. W. | Simple proof of security of the BB84 quantum key distribution protocol (2000) [100] |
| Bennett, C. H. | Experimental quantum cryptography (1992) [97] |

**Table 44**
Top publication fora in the quantum cryptography community.

| |
|---|
| Quantum Information Processing |
| Optics Communications |
| IEEE Photonics Technology Letters |
| International Journal of Theoretical Physics |
| Quantum Science and Technology |

**Table 45**
Most-cited authors (top five) in the quantum cryptography community.

| Author | Citations |
|---|---|
| Brassard, Gilles | 997 |
| Bennett, Charles H. | 866 |
| Crepeau, Claude | 689 |
| Salvail, Louis | 332 |
| Lo, Hoi-Kwong | 194 |

**Table 46**
Most-cited countries (top five) in the quantum cryptography community.

| Country | Citations |
|---|---|
| China | 4152 |
| United States | 3280 |
| Canada | 2931 |
| Switzerland | 1109 |
| United Kingdom | 511 |

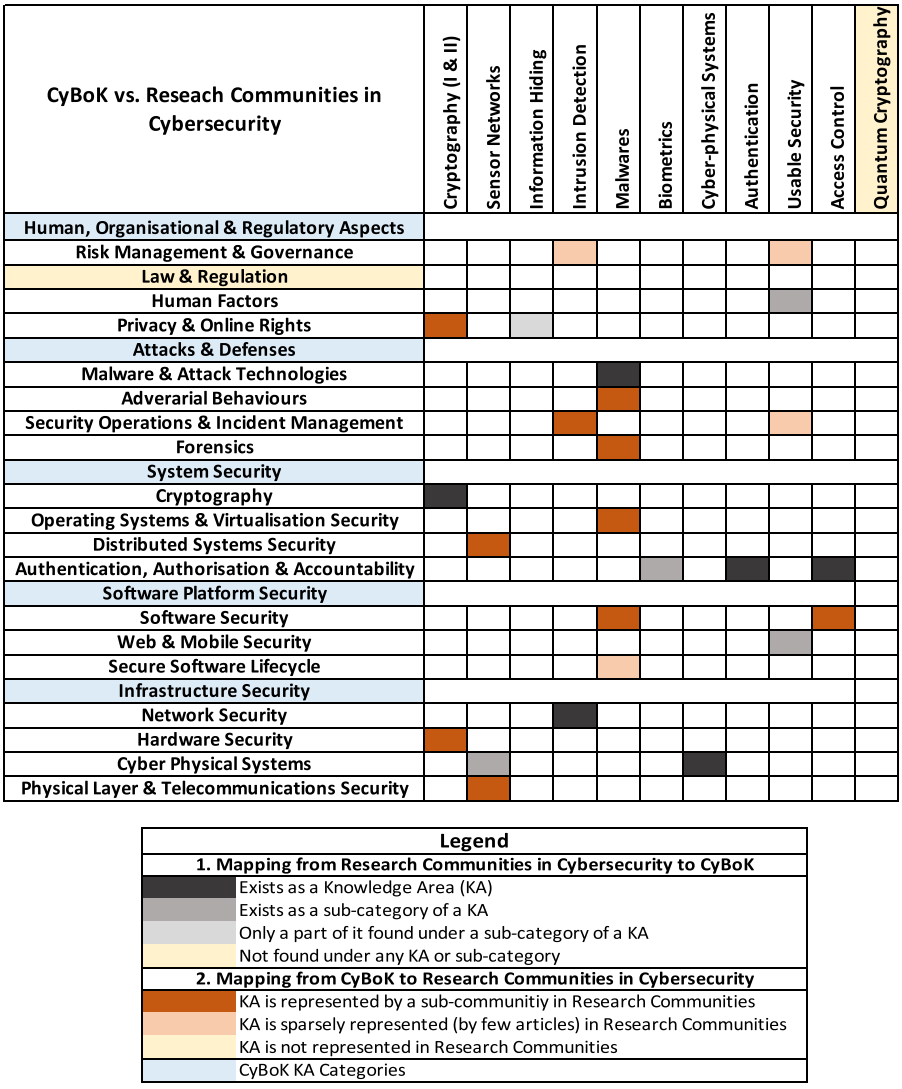


**Fig. 28.** Comparison matrix between CyBoK and our Communities in cyber security.

of inputs from academia and practitioners rather than targeting the research community. Nevertheless, there are many interesting commonalities between the knowledge areas of CyBOK and the researcher communities of the present work. The similarities and differences are discussed in greater detail below.

Another recently published study by Aniqua Baset and Tamara Denning [102] used Latent Dirichlet Allocation on a corpus of scientific articles to detect research areas, which were then clustered into topic categories to be further analyzed. It differs from the current work in terms of volume (3062 papers as compared to the 59,782 analyzed in the present work), time (1980–2015 as compared to 1949–2020 for the present work), its focus on the abstract topics rather than the researcher communities, and the employed analysis methods. However, the presented topics have similarities with the communities discovered in the present work. As with CyBOK, [102] and the present work are compared below.

In addition to the two aforementioned works spanning the whole topic of cyber security, there exist a number of literature reviews focusing on specific subtopics, such as cross-site scripting [103], information security management [104], security awareness [105], security and privacy in health [106], cloud computing risk [107], information security policy compliance [108], cyber situational awareness [109], digital forensics [110], phishing [111], threat modeling [112], and security requirements engineering [113].

There is also work employing similar methods as the current, but targeting other research areas. An example of such work is [114], where a similar automated approach for collecting and analyzing abstract and citation data was used.

### 5.1. Comparison to CyBoK

Since CyBoK [101] aims to identify the top knowledge areas (KAs) of cyber security, it constitutes a relevant object of comparison.

In Fig. 28, a comparison matrix between CyBoK and our work is presented. While there is a large overlap between the CyBOK knowledge areas and the researcher communities identified in the current work, there are some notable differences.

The **quantum cryptography** community is not found under any CyBOK knowledge area or knowledge area sub-category. Furthermore, **information hiding** and **biometrics** constitute significant research communities, but features less prominently in CyBOK. It is also noteworthy that **cryptography** is the overwhelmingly dominant research community, but does not appear to hold a similar position in CyBOK.

As compared to CyBOK, precious little research was identified in the fields of *laws and regulation*, as this topic did not even qualify for a sub-community. Additional CyBoK KAs that are only

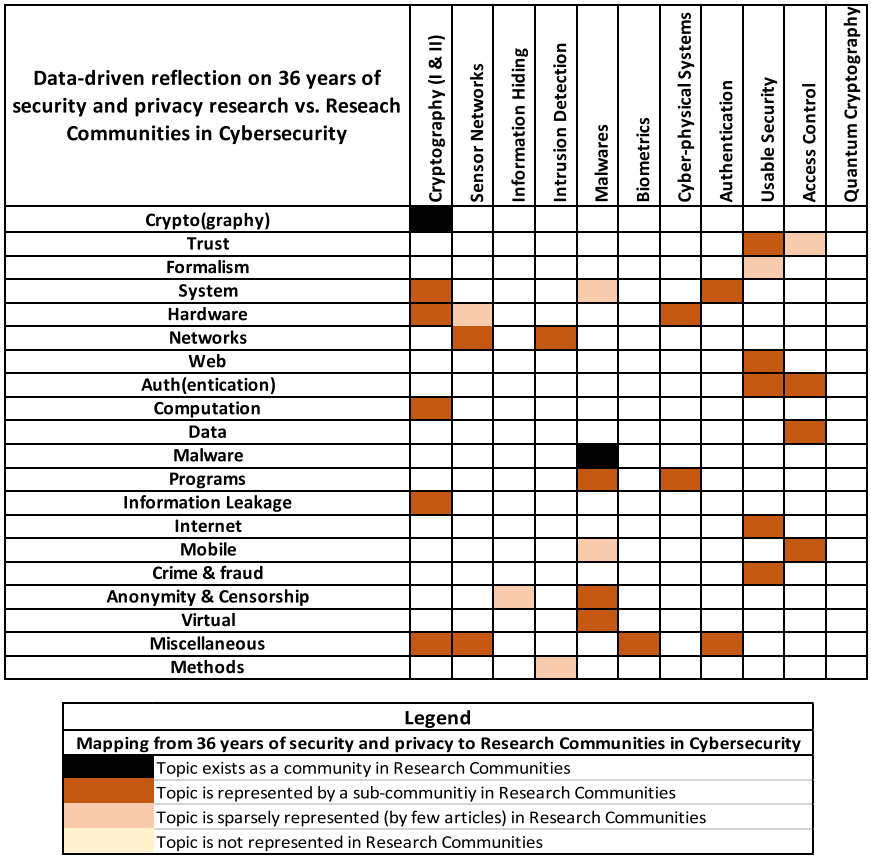


**Fig. 29.** Comparison matrix between Baset and Denning [102] and here reported communities in cyber security.

sparsely represented by the detected communities include *risk management and governance* and *security software lifecycle*.

Finally, the mid-sized **malwares** community covers aspects of several CyBOK knowledge areas, including *malware & attack technologies*, *adversarial behaviors*, *forensics*, *operating systems & virtualization security*, and *software security*, thus indicating another difference in emphasis between the research communities and CyBOK.

### 5.2. Comparison to Baset and Denning

As mentioned above, Baset and Denning [102] uses Latent Dirichlet Allocation to identify the topics in security and privacy research. In their article, 95 research topics were identified and categorized in 20 topic categories.

In Fig. 29, a comparison matrix between Baset and Denning [102] and our work is presented. The matrix demonstrates that all of Baset and Dennings' topics are either fully or partially represented by at least one sub-community. Considering the coverage of Baset and Denning's topics, we note that quantum cryptography was not present in any of their topic categories, which was also the case for CyBOK [101]. There is perfect alignment between the present and Baset and Denning's work for the "crypto" and "malware" topic categories. Considering the researcher communities' coverage of Baset and Denning's topics, the "formalism" and "methods" topics are the least represented by the communities detected on our work. It appears that these topics are distributed over many different communities.

It is not surprising that there are differences between the present and Baset and Denning's work. Latent Dirichlet Allocation considers the textual content of articles, while community detection is focused on the authors of those articles. One benefit of Latent Dirichlet Allocation is precisely the ability to abstract from the research process and organization, solely considering the produced results. Ideally, this approach would produce something similar to CyBOK, which also focuses on the abstract subject areas.

The citation relationships between researchers as presented in the present work is complementary to that of CyBOK and Baset and Denning. It provides information on the influence of one field on another, on the evolution of ideas, as well as on occasional topically inexplicable researcher behavior, such as why similar sub-communities sometimes maintain a distance. It also provides information on the geographical and organizational influence on different fields.

## 6. Discussion

There are a number of potential objections to the reliability and validity of the results presented in this article. That we only included the 59,782 most cited of the 320,907 articles might affect the results of the study. However, most of the omitted articles have less than a single digit number of citations and are therefore arguably unlikely to affect the community detection procedure.

Another threat to the validity of this study may be that older articles have received more citations than newer, simply because time has provided them more possibilities to be cited. This bias emphasizes older research over newer, and may thus also emphasize old research communities over newer ones. Time-normalizing citation counts is, however, not trivial, as citations are not necessarily a linear function of time — some articles continue to be cited long after publishing, while others do not, for instance. However, the results section provides plots of the annual article count per sub-community. These plots, such as

Fig. 3, inform about the relative importance of sub-communities at different points in time, allowing an unbiased determination of the research emphasis in recent years.

There is also the question of where the line is drawn for what constitutes a sub-community. We have defined it by a lower limit to the number of included authors. It would be possible to use other or additional criteria, such as the total number of citations.

Finally, the selection of Scopus as the (single) source of data has surely affected the results in terms of completeness. However, in addition to its broad coverage, Scopus also provides the application programming interface access which was required for this study.

## 7. Summary

By analyzing the most-cited scientific articles of 98,373 authors in the cyber security and information security domains, we were able to detect 12 research communities and sort them based on their current activity level: cryptography (I & II), sensor networks, information hiding, intrusion detection, malwares, biometrics, cyber–physical systems, authentication, usable security, access control, and quantum cryptography.

For each of these communities, we presented, among others, an overview of their topics, a discussion on their evolution over time, the sub-communities involved, and the most-cited articles.

As compared to related work aiming to represent both academia and practitioners, the presented research communities appear to place a greater emphasis on cryptography, quantum cryptography, information hiding and biometrics, at the expense of laws and regulation, risk management and governance, and security software lifecycle.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


[1] T.S. Kuhn, The Structure of Scientific Revolutions, University of Chicago Press, 1970.

[2] Michael Gusenbauer, Neal R Haddaway, Which academic search systems are suitable for systematic reviews or meta-analyses? evaluating retrieval qualities of google scholar, pubmed, and 26 other resources, Research Synthesis Methods 11 (2019) 181 – 217.

[3] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) (2008) P10008.

[4] R.L. Rivest, A. Shamir, L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, Commun. ACM 21 (2) (1978) 120–126, http://dx.doi.org/10.1145/359340.359342, URL https://doi-org.focus.lib.kth.se/10.1145/359340.359342.

[5] D. Boneh, M. Franklin, Identity-based encryption from the weil pairing, in: J. Kilian (Ed.), Advances in Cryptology — CRYPTO 2001, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, pp. 213–229.

[6] W. Diffie, M. Hellman, New directions in cryptography, IEEE Trans. Inform. Theory 22 (6) (2006) 644–654, http://dx.doi.org/10.1109/TIT.1976.1055638, URL https://doi-org.focus.lib.kth.se/10.1109/TIT.1976.1055638.

[7] A.J. Menezes, S.A. Vanstone, P.C.V. Oorschot, Handbook of Applied Cryptography, first ed., CRC Press, Inc., USA, 1996.

[8] A. Shamir, How to share a secret, Commun. ACM 22 (11) (1979) 612–613, http://dx.doi.org/10.1145/359168.359176, URL https://doi-org.focus.lib.kth.se/10.1145/359168.359176.

[9] P. Kocher, J. Jaffe, B. Jun, Differential power analysis, in: M. Wiener (Ed.), Advances in Cryptology — CRYPTO' 99, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999, pp. 388–397.

[10] A. Shamir, Identity-based cryptosystems and signature schemes, in: Proceedings of CRYPTO 84 on Advances in Cryptology, Springer-Verlag, Berlin, Heidelberg, 1985, pp. 47–53.

[11] T. Elgamal, A public key cryptosystem and a signature scheme based on discrete logarithms, IEEE Trans. Inform. Theory 31 (4) (1985) 469–472, http://dx.doi.org/10.1109/TIT.1985.1057074.

[12] M. Bellare, P. Rogaway, Random oracles are practical: A paradigm for designing efficient protocols, in: Proceedings of the 1st ACM Conference on Computer and Communications Security, CCS '93, Association for Computing Machinery, New York, NY, USA, 1993, pp. 62–73, http://dx.doi.org/10.1145/168588.168596, URL https://doi-org.focus.lib.kth.se/10.1145/168588.168596.

[13] P. Paillier, Public-key cryptosystems based on composite degree residuosity classes, in: J. Stern (Ed.), Advances in Cryptology — EUROCRYPT '99, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999, pp. 223–238.

[14] S. Goldwasser, Probabilistic Encryption: Theory and Applications (Partial Information, Factoring, Pseudo Random Bit Generation) (Ph.D. thesis), University of California, Berkeley, 1984, aAI8512835.

[15] R. Canetti, Universally composable security: A new paradigm for cryptographic protocols, in: Proceedings 42nd IEEE Symposium on Foundations of Computer Science, 2001, pp. 136–145, http://dx.doi.org/10.1109/SFCS.2001.959888.

[16] H. Feistel, Cryptography and Computer Privacy, Scientific American, 1973, URL https://books.google.se/books?id=0iYEzQEACAAJ.

[17] J. Daemen, V. Rijmen, The Design of Rijndael, Springer-Verlag, Berlin, Heidelberg, 2002.

[18] M. Matsui, Linear cryptanalysis method for DES cipher, in: Workshop on the Theory and Application of Cryptographic Techniques on Advances in Cryptology, EUROCRYPT '93, Springer-Verlag, Berlin, Heidelberg, 1994, pp. 386–397.

[19] D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, B. Sunar, Trojan detection using IC fingerprinting, in: Proceedings of the 2007 IEEE Symposium on Security and Privacy, SP '07, IEEE Computer Society, USA, 2007, pp. 296–310, http://dx.doi.org/10.1109/SP.2007.36.

[20] M. Tehranipoor, F. Koushanfar, A survey of hardware trojan taxonomy and detection, IEEE Des. Test Comput. 27 (1) (2010) 10–25, http://dx.doi.org/10.1109/MDT.2010.7.

[21] G.E. Suh, S. Devadas, Physical unclonable functions for device authentication and secret key generation, in: Proceedings of the 44th Annual Design Automation Conference, DAC '07, Association for Computing Machinery, New York, NY, USA, 2007, pp. 9–14, http://dx.doi.org/10.1145/1278480.1278484.

[22] V.S. Miller, Use of elliptic curves in cryptography, in: Lecture Notes in Computer Sciences; 218 on Advances in Cryptology—CRYPTO 85, Springer-Verlag, Berlin, Heidelberg, 1986, pp. 417–426.

[23] N. Koblitz, Elliptic curve cryptosystems, Math. Comp. 48 (177) (1987) 203–209.

[24] C. Gentry, Fully homomorphic encryption using ideal lattices, in: Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing, STOC '09, Association for Computing Machinery, New York, NY, USA, 2009, pp. 169–178, http://dx.doi.org/10.1145/1536414.1536440.

[25] D. Boneh, G. Di Crescenzo, R. Ostrovsky, G. Persiano, Public key encryption with keyword search, in: C. Cachin, J.L. Camenisch (Eds.), Advances in Cryptology - EUROCRYPT 2004, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 506–522.

[26] L. Sweeney, K-Anonymity: A model for protecting privacy, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 10 (5) (2002) 557–570, http://dx.doi.org/10.1142/S0218488502001648.

[27] J. Bethencourt, A. Sahai, B. Waters, Ciphertext-policy attribute-based encryption, in: Proceedings of the 2007 IEEE Symposium on Security and Privacy, SP '07, IEEE Computer Society, USA, 2007, pp. 321–334, http://dx.doi.org/10.1109/SP.2007.11.

[28] A. Benefit, The Evolution of Wireless Sensor Networks, Silicon Laboratories Inc., 2013, URL https://www.silabs.com/documents/public/white-papers/evolution-of-wireless-sensor-networks.pdf.

[29] Vipul Gupta, M. Millard, S. Fung, Y. Zhu, N. Gura, H. Eberle, S.C. Shantz, Sizzle: A standards-based end-to-end security architecture for the embedded Internet, in: Third IEEE International Conference on Pervasive Computing and Communications, 2005, pp. 247–256, http://dx.doi.org/10.1109/PERCOM.2005.41.

[30] M. Bloch, J. Barros, M.R.D. Rodrigues, S.W. McLaughlin, Wireless information-theoretic security, IEEE Trans. Inform. Theory 54 (6) (2008) 2515–2534, http://dx.doi.org/10.1109/TIT.2008.921908.

[31] L. Eschenauer, V.D. Gligor, A key-management scheme for distributed sensor networks, in: Proceedings of the 9th ACM Conference on Computer and Communications Security, CCS '02, ACM, New York, NY, USA, 2002, pp. 41–47, http://dx.doi.org/10.1145/586110.586117, URL http://doi.acm.org/10.1145/586110.586117.

[32] M. Raya, J.-P. Hubaux, Securing vehicular Ad Hoc networks, J. Comput. Secur. 15 (1) (2007) 39–68.

[33] A. Perrig, R. Szewczyk, J.D. Tygar, V. Wen, D.E. Culler, SPINS: Security protocols for sensor networks, Wirel. Netw. 8 (5) (2002) 521–534, http://dx.doi.org/10.1023/A:1016598314198.

[34] C. Karlof, D. Wagner, Secure routing in wireless sensor networks: Attacks and countermeasures, in: Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, 2003, 2003, pp. 113–127, http://dx.doi.org/10.1109/SNPA.2003.1203362.

[35] H. Chan, A. Perrig, D. Song, Random key predistribution schemes for sensor networks, in: 2003 Symposium on Security and Privacy, 2003, 2003, pp. 197–213, http://dx.doi.org/10.1109/SECPRI.2003.1199337.

[36] J.R. Douceur, The sybil attack, in: P. Druschel, F. Kaashoek, A. Rowstron (Eds.), Peer-to-Peer Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2002, pp. 251–260.

[37] C.E. Shannon, Communication theory of secrecy systems, Bell Syst. Tech. J. 28 (4) (1949) 656–715, http://dx.doi.org/10.1002/j.1538-7305.1949.tb00928.x.

[38] F. Hartung, M. Kutter, Multimedia watermarking techniques, Proc. IEEE 87 (7) (1999) 1079–1107, http://dx.doi.org/10.1109/5.771066.

[39] G. Chen, Y. Mao, C.K. Chui, A symmetric image encryption scheme based on 3D chaotic cat maps, Chaos Solitons Fractals 21 (3) (2004) 749–761, http://dx.doi.org/10.1016/j.chaos.2003.12.022, URL http://www.sciencedirect.com/science/article/pii/S0960077903006672.

[40] Jun Tian, Reversible data embedding using a difference expansion, IEEE Trans. Circuits Syst. Video Technol. 13 (8) (2003) 890–896, http://dx.doi.org/10.1109/TCSVT.2003.815962.

[41] I.J. Cox, J. Kilian, F.T. Leighton, T. Shamoon, Secure spread spectrum watermarking for multimedia, IEEE Trans. Image Process. 6 12 (1997) 1673–1687.

[42] F.A.P. Petitcolas, R.J. Anderson, M.G. Kuhn, Information hiding-a survey, Proc. IEEE 87 (7) (1999) 1062–1078, http://dx.doi.org/10.1109/5.771065.

[43] S. Forrest, S.A. Hofmeyr, A. Somayaji, T.A. Longstaff, A sense of self for unix processes, in: Proceedings of the 1996 IEEE Symposium on Security and Privacy, SP '96, IEEE Computer Society, USA, 1996, p. 120.

[44] P. Ammann, D. Wijesekera, S. Kaushik, Scalable, graph-based network vulnerability analysis, in: Proceedings of the 9th ACM Conference on Computer and Communications Security, ACM, 2002, pp. 217–224.

[45] D.E. Denning, An intrusion-detection model, IEEE Trans. Softw. Eng. 13 (2) (1987) 222–232, http://dx.doi.org/10.1109/TSE.1987.232894.

[46] S. Savage, D. Wetherall, A. Karlin, T. Anderson, Practical network support for IP traceback, in: ACM SIGCOMM Computer Communication Review, Vol. 30, ACM, 2000, pp. 295–306.

[47] S. Dharmapurikar, P. Krishnamurthy, T.S. Sproull, J.W. Lockwood, Deep packet inspection using parallel bloom filters, IEEE Micro 24 (1) (2004) 52–61, http://dx.doi.org/10.1109/MM.2004.1268997.

[48] K. Lakkaraju, W. Yurcik, A.J. Lee, NVisionIP: Netflow visualizations of system state for security situational awareness, in: Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security, VizSEC/DMSEC '04, Association for Computing Machinery, New York, NY, USA, 2004, pp. 65–72, http://dx.doi.org/10.1145/1029208.1029219.

[49] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, M. Rajarajan, Review: A survey of intrusion detection techniques in cloud, J. Netw. Comput. Appl. 36 (1) (2013) 42–57, http://dx.doi.org/10.1016/j.jnca.2012.05.003.

[50] Rodrigo Braga, Edjard Mota, Alexandre Passito, Lightweight DDoS flooding attack detection using NOX/OpenFlow, in: Proceedings of the 2010 IEEE 35th Conference on Local Computer Networks, LCN '10, IEEE Computer Society, USA, 2010, pp. 408–415, http://dx.doi.org/10.1109/LCN.2010.5735752.

[51] S. Shin, V. Yegneswaran, P. Porras, G. Gu, AVANT-GUARD: Scalable and vigilant switch flow management in software-defined networks, in: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, CCS '13, Association for Computing Machinery, New York, NY, USA, 2013, pp. 413–424, http://dx.doi.org/10.1145/2508859.2516684.

[52] J. McHugh, Testing intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln laboratory, ACM Trans. Inf. Syst. Secur. 3 (4) (2000) 262–294, http://dx.doi.org/10.1145/382912.382923.

[53] O. Sheyner, J. Haines, S. Jha, R. Lippmann, J.M. Wing, Automated generation and analysis of attack graphs, in: Proceedings 2002 IEEE Symposium on Security and Privacy, 2002, pp. 273–284, http://dx.doi.org/10.1109/SECPRI.2002.1004377.

[54] R.P. Lippmann, D.J. Fried, I. Graf, J.W. Haines, K.R. Kendall, D. McClung, D. Weber, S.E. Webster, D. Wyschogrod, R.K. Cunningham, M.A. Zissman, Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation, in: Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00, Vol. 2, 2000, pp. 12–26, http://dx.doi.org/10.1109/DISCEX.2000.821506.

[55] D.E. Denning, P.J. Denning, Certification of programs for secure information flow, Commun. ACM 20 (7) (1977) 504–513.

[56] J.A. Goguen, J. Meseguer, Security policies and security models, in: 1982 IEEE Symposium on Security and Privacy, IEEE, 1982, p. 11.

[57] F. Cohen, Computer viruses: Theory and experiments, Comput. Secur. 6 (1) (1987) 22–35.

[58] J.O. Kephart, S.R. White, Directed-graph epidemiological models of computer viruses, in: Computation: The Micro and the Macro View, World Scientific, 1992, pp. 71–102.

[59] W. Enck, P. Gilbert, S. Han, V. Tendulkar, B.-G. Chun, L.P. Cox, J. Jung, P. McDaniel, A.N. Sheth, TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones, ACM Trans. Comput. Syst. 32 (2) (2014) 5.

[60] M.G. Schultz, E. Eskin, F. Zadok, S.J. Stolfo, Data mining methods for detection of new malicious executables, in: Proceedings 2001 IEEE Symposium on Security and Privacy. S&P 2001, IEEE, 2000, pp. 38–49.

[61] Y. Zhou, X. Jiang, Dissecting android malware: Characterization and evolution, in: 2012 IEEE Symposium on Security and Privacy, IEEE, 2012, pp. 95–109.

[62] L. Huang, A.D. Joseph, B. Nelson, B.I.P. Rubinstein, J.D. Tygar, Adversarial machine learning, in: Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence, AISec '11, Association for Computing Machinery, New York, NY, USA, 2011, pp. 43–58, http://dx.doi.org/10.1145/2046684.2046692, URL https://doi-org.focus.lib.kth.se/10.1145/2046684.2046692.

[63] A. Sabelfeld, A.C. Myers, Language-based information-flow security, IEEE J. Sel. Areas Commun. 21 (1) (2003) 5–19, http://dx.doi.org/10.1109/JSAC.2002.806121.

[64] A.P. Felt, E. Chin, S. Hanna, D. Song, D. Wagner, Android permissions demystified, in: Proceedings of the 18th ACM Conference on Computer and Communications Security, CCS '11, Association for Computing Machinery, New York, NY, USA, 2011, pp. 627–638, http://dx.doi.org/10.1145/2046707.2046779.

[65] W. Enck, M. Ongtang, P. McDaniel, On lightweight mobile phone application certification, in: Proceedings of the 16th ACM Conference on Computer and Communications Security, ACM, 2009, pp. 235–245.

[66] U. Uludag, S. Pankanti, S. Prabhakar, A.K. Jain, Biometric cryptosystems: Issues and challenges, Proc. IEEE 92 (6) (2004) 948–960, http://dx.doi.org/10.1109/JPROC.2004.827372.

[67] G.I. Davida, Y. Frankel, B.J. Matt, On enabling secure applications through off-line biometric identification, in: Proceedings. 1998 IEEE Symposium on Security and Privacy (Cat. No.98CB36186), 1998, pp. 148–157, http://dx.doi.org/10.1109/SECPRI.1998.674831.

[68] J. Schiff, M. Meingast, D.K. Mulligan, S. Sastry, K. Goldberg, Respectful cameras: Detecting visual markers in real-time to address privacy concerns, in: Protecting Privacy in Video Surveillance, Springer London, London, 2009, pp. 65–89.

[69] R. Joyce, G. Gupta, Identity authentication based on keystroke latencies, Commun. ACM 33 (2) (1990) 168–176, http://dx.doi.org/10.1145/75577.75582.

[70] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Trans. Circuits Syst. Video Technol. 14 (1) (2004) 4–20, http://dx.doi.org/10.1109/TCSVT.2003.818349.

[71] A. Juels, M. Wattenberg, A fuzzy commitment scheme, in: Proceedings of the 6th ACM Conference on Computer and Communications Security, CCS '99, ACM, New York, NY, USA, 1999, pp. 28–36, http://dx.doi.org/10.1145/319709.319714, URL http://doi.acm.org/10.1145/319709.319714.

[72] N.K. Ratha, J.H. Connell, R.M. Bolle, Enhancing security and privacy in biometrics-based authentication systems, IBM Syst. J. 40 (3) (2001) 614–634.

[73] N.K. Ratha, S. Chikkerur, J.H. Connell, R.M. Bolle, Generating cancelable fingerprint templates, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 561–572.

[74] S. Sridhar, A. Hahn, M. Govindarasu, Cyber–physical system security for the electric power grid, Proc. IEEE 100 (1) (2012) 210–224, http://dx.doi.org/10.1109/JPROC.2011.2165269.

[75] Y. Liu, P. Ning, M.K. Reiter, False data injection attacks against state estimation in electric power grids, in: Proceedings of the 16th ACM Conference on Computer and Communications Security, CCS '09, ACM, New York, NY, USA, 2009, pp. 21–32, http://dx.doi.org/10.1145/1653662.1653666, URL http://doi.acm.org/10.1145/1653662.1653666.

[76] F. Pasqualetti, F. Dörfler, F. Bullo, Attack detection and identification in cyber-physical systems, IEEE Trans. Automat. Control 58 (11) (2013) 2715–2729, http://dx.doi.org/10.1109/TAC.2013.2266831.

[77] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, et al., Experimental security analysis of a modern automobile, in: Security and Privacy, SP, 2010 IEEE Symposium on, IEEE, 2010, pp. 447–462.

[78] O. Kosut, L. Jia, R.J. Thomas, L. Tong, Malicious data attacks on the smart grid, IEEE Trans. Smart Grid 2 (4) (2011) 645–658, http://dx.doi.org/10.1109/TSG.2011.2163807.

[79] M.L. Das, Two-factor user authentication in wireless sensor networks, IEEE Trans. Wirel. Commun. 8 (3) (2009) 1086–1090, http://dx.doi.org/10.1109/TWC.2008.080128.

[80] L. Lamport, Password authentication with insecure communication, Commun. ACM 24 (11) (1981) 770–772, http://dx.doi.org/10.1145/358790.358797.

[81] D. Dolev, A. Yao, On the security of public key protocols, IEEE Trans. Inform. Theory 29 (2) (1983) 198–208.

[82] T.S. Messerges, E.A. Dabbish, R.H. Sloan, Examining smart-card security under the threat of power analysis attacks, IEEE Trans. Comput. 51 (5) (2002) 541–552.
[83] M. Burrows, M. Abadi, R. Needham, A logic of authentication, ACM Trans. Comput. Syst. 8 (1) (1990) 18–36, http://dx.doi.org/10.1145/77648.77649, URL https://doi-org.focus.lib.kth.se/10.1145/77648.77649.
[84] R.W. Rogers, A protection motivation theory of fear appeals and attitude change1, J. Psychol. 91 (1) (1975) 93–114, http://dx.doi.org/10.1080/00223980.1975.9915803, PMID: 28136248.
[85] D.W. Straub, R.J. Welke, Coping with systems risk: Security planning models for management decision making, MIS Q. (1998) 441–469.
[86] R. Dhamija, J.D. Tygar, M. Hearst, Why phishing works, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2006, pp. 581–590.
[87] T. Herath, H.R. Rao, Protection motivation and deterrence: A framework for security policy compliance in organisations, Eur. J. Inf. Syst. 18 (2) (2009) 106–125, http://dx.doi.org/10.1057/ejis.2009.6.
[88] B. Bulgurcu, H. Cavusoglu, I. Benbasat, Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness, MIS Q. 34 (2010) 523–548, http://dx.doi.org/10.2307/25750690.
[89] A.C. Johnston, M. Warkentin, Fear appeals and information security behaviors: An empirical study, MIS Q. 34 (3) (2010) 549–566.
[90] R.S. Sandhu, E.J. Coyne, H.L. Feinstein, C.E. Youman, Role-based access control models, Computer 29 (2) (1996) 38–47, http://dx.doi.org/10.1109/2.485845, URL https://doi-org.focus.lib.kth.se/10.1109/2.485845.
[91] M. Blaze, J. Feigenbaum, J. Lacy, Decentralized trust management, in: Proceedings of the 1996 IEEE Conference on Security and Privacy, SP'96, IEEE Computer Society, Washington, DC, USA, 1996, pp. 164–173, URL http://dl.acm.org.focus.lib.kth.se/citation.cfm?id=1947337.1947361.
[92] G. Sindre, A.L. Opdahl, Eliciting security requirements by misuse cases, in: Proceedings 37th International Conference on Technology of Object-Oriented Languages and Systems. TOOLS-Pacific 2000, 2000, pp. 120–131, http://dx.doi.org/10.1109/TOOLS.2000.891363.
[93] D.F. Ferraiolo, R. Sandhu, S. Gavrila, D.R. Kuhn, R. Chandramouli, Proposed NIST standard for role-based access control, ACM Trans. Inf. Syst. Secur. 4 (3) (2001) 224–274, http://dx.doi.org/10.1145/501978.501980, URL http://doi.acm.org.focus.lib.kth.se/10.1145/501978.501980.
[94] E. Bertino, P.A. Bonatti, E. Ferrari, TRBAC: A temporal role-based access control model, ACM Trans. Inf. Syst. Secur. 4 (3) (2001) 191–233, http://dx.doi.org/10.1145/501978.501979, URL http://doi.acm.org.focus.lib.kth.se/10.1145/501978.501979.
[95] C.H. Bennett, Quantum cryptography using any two nonorthogonal states, Phys. Rev. Lett. 68 (1992) 3121–3124, http://dx.doi.org/10.1103/PhysRevLett.68.3121, URL https://link.aps.org/doi/10.1103/PhysRevLett.68.3121.
[96] F.-G. Deng, G.L. Long, Secure direct communication with a quantum one-time pad, Phys. Rev. A 69 (5) (2004) http://dx.doi.org/10.1103/physreva.69.052319, URL http://dx.doi.org/10.1103/PhysRevA.69.052319.
[97] C.H. Bennett, F. Bessette, G. Brassard, L. Salvail, J. Smolin, Experimental quantum cryptography, J. Cryptol. 5 (1) (1992) 3–28, http://dx.doi.org/10.1007/BF00191318.
[98] C. Bennett, D. Zekrifa, Quantum cryptography: Public key distribution and coin tossing, in: Proceedings of IEEE international Conference on Computers, Systems and Signal Processing, 2014, pp. 175–179.
[99] N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, Quantum cryptography, Rev. Modern Phys. 74 (2002) 145–195, http://dx.doi.org/10.1103/RevModPhys.74.145, URL https://link.aps.org/doi/10.1103/RevModPhys.74.145.
[100] P.W. Shor, J. Preskill, Simple proof of security of the BB84 quantum key distribution protocol, Phys. Rev. Lett. 85 (2000) 441–444, http://dx.doi.org/10.1103/PhysRevLett.85.441, URL https://link.aps.org/doi/10.1103/PhysRevLett.85.441.
[101] J. Hallett, R. Larson, A. Rashid, Mirror, mirror, on the wall: What are we teaching them all? Characterising the focus of cybersecurity curricular frameworks, in: 2018 USENIX Workshop on Advances in Security Education, ASE 18, USENIX Association, Baltimore, MD, 2018, URL https://www.usenix.org/conference/ase18/presentation/hallett.
[102] A. Baset, T. Denning, A data-driven reflection on 36 years of security and privacy research, in: 12th USENIX Workshop on Cyber Security Experimentation and Test, CSET 19, USENIX Association, Santa Clara, CA, 2019, URL https://www.usenix.org/conference/cset19/presentation/baset.
[103] I. Hydara, A.B.M. Sultan, H. Zulzalil, N. Admodisastro, Current state of research on cross-site scripting (XSS)–A systematic literature review, Inf. Softw. Technol. 58 (2015) 170–186.
[104] Z.A. Soomro, M.H. Shah, J. Ahmed, Information security management needs more holistic approach: A literature review, Int. J. Inf. Manage. 36 (2) (2016) 215–225.
[105] B. Lebek, J. Uffen, M.H. Breitner, M. Neumann, B. Hohler, Employees' information security awareness and behavior: A literature review, in: System Sciences, HICSS, 2013 46th Hawaii International Conference on, IEEE, 2013, pp. 2978–2987.
[106] J.L. Fernández-Alemán, I.C. Señor, P.Á.O. Lozoya, A. Toval, Security and privacy in electronic health records: A systematic literature review, J. Biomed. Inform. 46 (3) (2013) 541–562.
[107] R. Latif, H. Abbas, S. Assar, Q. Ali, Cloud computing risk assessment: A systematic literature review, in: Future Information Technology, Springer, 2014, pp. 285–295.
[108] T. Sommestad, J. Hallberg, K. Lundholm, J. Bengtsson, Variables influencing information security policy compliance: A systematic review of quantitative studies, Inf. Manage. Comput. Secur. 22 (1) (2014) 42–75.
[109] U. Franke, J. Brynielsson, Cyber situational awareness–a systematic review of the literature, Comput. Secur. 46 (2014) 18–31.
[110] S. Alharbi, J. Weber-Jahnke, I. Traore, The proactive and reactive digital forensics investigation process: A systematic literature review, in: International Conference on Information Security and Assurance, Springer, 2011, pp. 87–100.
[111] S. Das, A. Kim, Z. Tingle, C. Nippert-Eng, All about phishing: Exploring user research through a systematic literature review, 2019, arXiv preprint arXiv:1908.05897.
[112] W. Xiong, R. Lagerström, Threat modeling–A systematic literature review, Comput. Secur. (2019).
[113] D. Mellado, C. Blanco, L.E. Sánchez, E. Fernández-Medina, A systematic review of security requirements engineering, Comput. Stand. Interfaces 32 (4) (2010) 153–165.
[114] M.V. Mäntylä, D. Graziotin, M. Kuutila, The evolution of sentiment analysis—A review of research topics, venues, and top cited papers, Comp. Sci. Rev. 27 (2018) 16–32, http://dx.doi.org/10.1016/j.cosrev.2017.10.002.

**Sotirios Katsikeas** is a Ph.D. student at the School of Electrical and Computer Engineering, KTH Royal Institute of Technology, Sweden. His main topic of research is cybersecurity modeling and attack simulations. He received his Diploma (5-year program) in Electrical and Computer Engineering from Technical University of Crete, Greece and his M.Sc. from the KTH Royal Institute of Technology in 2018.

**Pontus Johnson** is a professor at the Royal Institute of Technology (KTH) in Stockholm, Sweden. His research interests mainly lie in the area of cyber security and the analysis of architectural models of computer networks — in particular simulating cyber attacks on such networks. Pontus supervises a number of Ph.D. students and holds courses on Ethical Hacking. He is, since 2020, the director of the Center for Cyber Defense and Information Security at KTH. He received his M.Sc. from the Lund Institute of Technology in 1997 and his Ph.D. and Docent titles from the Royal Institute of Technology in 2002 and 2007. He was appointed professor in 2009.

**Mathias Ekstedt** is professor of Industrial Information and Control Systems at KTH Royal Institute of Technology in Stockholm, Sweden. He received his M.Sc. and Ph.D. from the Royal Institute of Technology in 1999 and 2004 respectively. Mathias' research interests include information and cyber security based on software and systems architecture modeling and analysis. In particular the research is applied in the power industry and information systems related to physical process monitoring and control, such as SCADA and Industrial Control Systems.

**Robert Lagerström** is an associate professor at the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Sweden. His research is focused on enterprise security, threat modeling, attack simulations, vehicle security, and viable cities. Robert is a member of the Young Academy of Sweden and one of the founders of foreseeti.