

CyCoAnalysis: A Dataset of Cybersecurity Conference Metadata for Meta-Science and Policy Decisions

Marton Bognar
DistriNet, KU Leuven
Leuven, Belgium
marton.bognar@kuleuven.be

Lieven Desmet
DistriNet, KU Leuven
Leuven, Belgium
lieven.desmet@kuleuven.be

Frank Piessens
DistriNet, KU Leuven
Leuven, Belgium
frank.piessens@kuleuven.be

Abstract—A large share of high-impact cybersecurity research is published at conferences. Access to high-quality data about these conferences is crucial for meta-science, enabling research into how policy decisions influence the practices of researchers and vice versa, how changing trends in research are reflected in conference organization and submission requirements. Furthermore, such a dataset would enable more data-driven decision-making for conference organizers, bringing more transparency and potentially more homogeneous requirements. At this moment, however, a large share of the information about past conferences is difficult to access, scattered, and not available in a structured format.

To solve this challenge, we present the CyCoAnalysis (*Cybersecurity Conference Analysis*) dataset, collecting metadata on highly-ranked cybersecurity conferences from the past 20 years. We collect details about the conference organization, submission requirements, publication details, and more. In this paper, we describe the dataset, demonstrate overarching trends, and enumerate research directions that could benefit from our data. Our vision is to engage the community to use and contribute to this dataset, enlarging its scope in the future.

1. Introduction

Cybersecurity research is often published at conferences, representing some of the most prestigious venues in the field. Consequently, these conferences play an important role in career advancement and funding considerations for researchers who rely on them for publishing their results and networking. While the main conferences in the field have similar goals, they operate largely independently, diverging in their implemented policies and addressing similar challenges with different approaches. The decisions related to the content of the conferences are primarily made by the program chairs (usually supported by a steering committee, responsible for the long-term vision), with limited transparency. Importantly, the program chairs often change yearly, bringing constant change in the policies at the conferences. These policies are reported on the conference website in the form of a call for papers (CfP), along with other information such as the conference location and details of the proceedings.

Unfortunately, collecting past conference websites and distilling the relevant data in an easy-to-parse format is difficult. There are community initiatives (sometimes outdated or discontinued) on collecting data related to cybersecurity conferences [2], [5], [6], [12], [15], [20], but these often serve specific purposes, such as ranking conferences and authors, and are not community-driven. The two most closely related initiatives are sec-deadlines [17] and secartifacts [18], which both encourage community involvement. The first one collects upcoming submission deadlines for cybersecurity conferences (and journals), while the second facilitates and stores the outcomes of artifact evaluation.

Our contribution. In this paper, we argue that collecting and storing a broader set of historical conference data in an extensible format is crucial, and to this end, propose a new structured dataset, CyCoAnalysis. We believe that such a dataset will serve at least two purposes and target audiences. First, conference and workshop organizers will easily be able to get an overview of different policies that were proposed and tried out at past events, see their historical evolution, and even know which organizers to contact for a first-hand account. This should enable more informed decision-making for future events based on the observed trends and, potentially, their impact on the published papers. Second, we hope that our dataset will support a large variety of meta-research activities in cybersecurity. Using our data, researchers can investigate how our field has evolved over the years, how it has adopted insights from other scientific disciplines on peer-review and publication practices, and how it has changed in terms of growth and diversity of the community. Additionally, our data enables analysis into how submission policies can impact the behavior of authors in terms of, e.g., the topic of published papers, the issues discussed in them (such as ethical concerns), and the choice of the publication venue itself. The demand for such a dataset is clear, with many meta-science studies analyzing or discussing certain aspects of conference policies, such as mandatory data sharing statements [3], [6], [11], artifact evaluation policies [3], [6], [9], [10], [11], topics of interest [6], [7], and others [6], [7], [16], [19]. Our dataset covers all these aspects related to submission policies at the past 20 years of CORE A* and A conferences along with other metadata on the conferences, and is designed to be

extended by the community.

Outline. We start by describing our dataset in [Section 2](#), then present preliminary insights in [Section 3](#), and finally propose a number of future work and research directions in [Section 4](#), both for extending our dataset and using it in meta-science research.

Open science. Our dataset and related materials are available and open to contributions at <https://cycoanalysis.github.io>.

2. The CyCoAnalysis dataset

At the time of publication, our dataset covers the past 20 years (2005-2025) of CORE A* (IEEE S&P, USENIX SEC, ACM CCS, NDSS) and most CORE A (ACSAC, ASIA CCS, CSF, ESORICS, EuroS&P, and RAID) cybersecurity conference editions. All data was manually extracted from the conference websites and other sources, such as published proceedings, and adapted into a uniform format. Each edition is described in a YAML file with an associated schema, aiming for a balance between machine- and human readability. In this section, we briefly describe the most important types of metadata currently covered in the dataset and their sources.

Conference website. We list the website for each conference edition, together with a boolean flag indicating whether the website is available at the time of analysis. For certain conferences, finding the original URL turned out to be a significant challenge, as many (especially A-ranked) conferences are hosted on different domains each year (often institutional subpages), which are removed after the conference, and no up-to-date historical list exists. In these cases, we found the website by searching, e.g., public mailing lists and websites such as WikiCFP (which also does not consistently feature these conferences). For the unavailable websites, we also store a URL to a snapshot on the Internet Archive. Despite these efforts, we have not yet managed to recover all CFPs from A-ranked conferences. The challenges we encountered with finding and retrieving the conference websites highlight the importance of a centralized and publicly available dataset.

Program committee. We collect the program committee (PC) for all editions as listed on the website. This includes the full name and affiliation of the members, with the country additionally being listed in certain cases. We separately collect the program chair(s) and vice- or area-chairs.

Submission deadlines. For each submission cycle, we store all deadlines from the CFP. These usually include the submission date, acceptance notification, and camera-ready deadline. When the conference features revisions, a rebuttal period, or an early rejection round, these are also listed.

Topics. Most conference websites explicitly list the topics of interest for that year. Currently, we only collect the standard topics, not calls for specific paper types such as Systematization of Knowledge (SoK) or special themes.

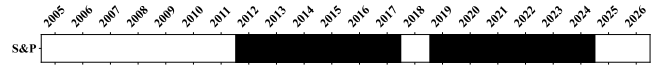


Figure 1. Marked (dark) editions of S&P are fully open access.

Peer review. We collect two responsibilities of the authors for the peer-review process: whether they need to anonymize their submissions, and whether conflicts of interest need to be declared. We also note whether the conference publishes a meta-review along with accepted papers.

Ethics, open science, and AI usage. Our dataset also contains information on whether the authors need to consider the ethical consequences of their research and how their research code and data will be shared. Furthermore, we record whether the conference has guidance or requirements on generative AI usage. Currently, we only store these features as binary flags, not considering whether they are mandatory or the depth of the relevant guidelines (cf. [Section 4](#)).

Proceedings. For each conference edition, we store the corresponding DBLP entry for the proceedings, which can enable further analysis on the papers themselves. Notably, the proceedings are often not linked from the conference website itself, and especially A-ranked conferences do not always use the same publisher.

Open access status. We review whether the proceedings for a given edition (*all* papers) can be accessed free of charge from the publisher’s website.

3. A glimpse into the data

In this section, we share preliminary insights and illustrate some overarching trends at the four A* conferences based on our dataset.

Open access status. USENIX SEC and NDSS have been open access for many years, while ACM (the publisher of CCS) has recently switched to a Gold Open Access model [1]. This model means that articles (also from older editions) can be accessed for free, but authors have to pay a mandatory publishing charge going forward. This leaves IEEE S&P as the last conference that is not fully open access, with [Figure 1](#) showing the closed-access editions in white.

Community growth. The number of published papers and PC members has been steadily growing. [Figure 2](#) shows the number of papers per conference edition, obtained through our collected DBLP reference for the proceedings, with a steep increase around 2020. Interestingly, CCS suffered a dip that year, which could be explained by the submission deadlines falling after the start of the global pandemic. [Figure 3](#) shows a similar trend in the number of PC members. To give an indication of the workload per PC member, [Figure 4](#) shows the ratio of *published* papers per PC member. Note that this representation of the workload is only indicative: it

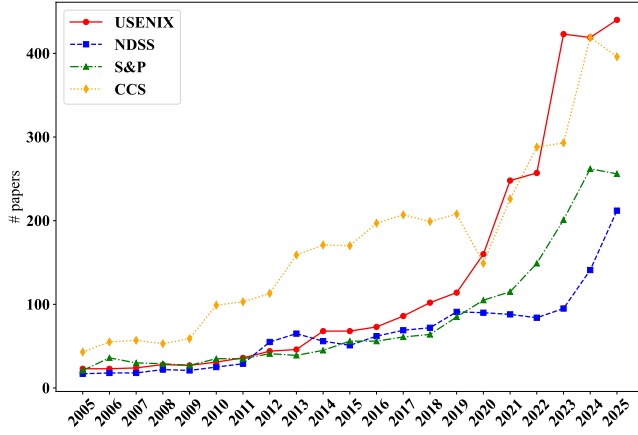


Figure 2. Number of published papers per conference edition.

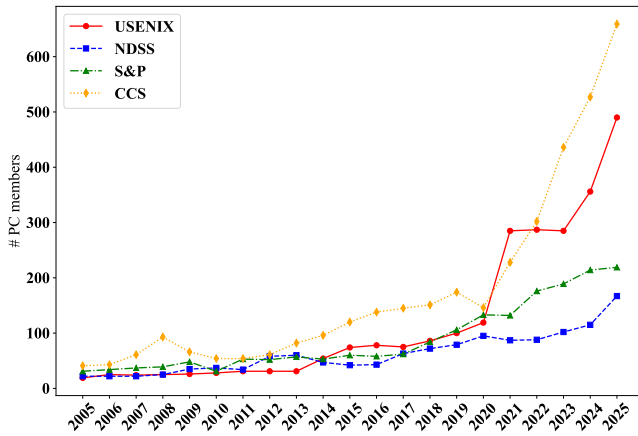


Figure 3. Number of PC members per conference edition.

might be skewed by different acceptance rates and uneven workloads within the PC, especially at CCS, where papers are submitted to and reviewed by separate conference tracks, which we do not consider here.

Geographical diversity. Certain conference websites list the country of individual PC members, allowing us to get a glimpse into the geographical diversity of PCs without enriching the data. By picking a few examples, we observe an initially very US-dominated field that is becoming a more global community. At CCS, more than 75% (33/43) of the PC was US-based in 2006, while this reduced to 34% by 2024 (182/527). NDSS used to be an even more US-centric conference, with 33/35 US-based PC members in 2009.

Peer-review policies. While today there is significant overlap in the policies related to the peer-review process, we found that their history differs significantly. Declaring conflicts of interest with the PC members first appeared at USENIX SEC in 2015, and was implemented by all four conferences two years later. Recently, guidelines on generative AI use were also adopted by all conferences in a span of less than two

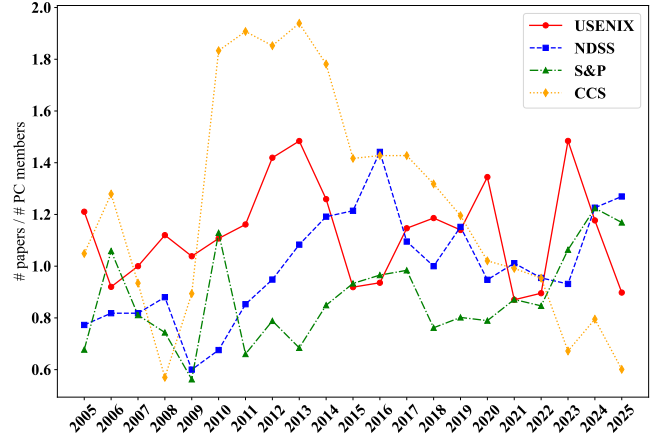


Figure 4. Rate of published papers and PC size per conference edition.

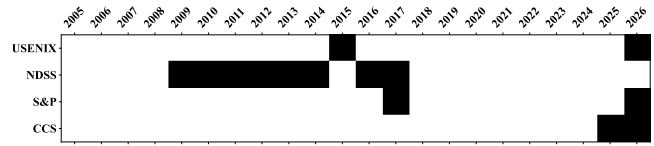


Figure 5. The marked (dark) conference editions required authors to register submission titles and abstracts before the submission deadline.

years. These stand in contrast to anonymizing submissions, which has been required since the starting date of our data at CCS (with the notable exceptions of 2010–2011) and all S&P editions, but was not required at USENIX SEC and NDSS until 2011. Since 2025, S&P additionally prohibits including CVE identifiers in submissions to preserve anonymity, a rule that so far has not been adopted elsewhere. S&P is also alone among A* conferences in publishing meta-reviews that summarize the reviewer comments for each accepted paper since 2024. Ethics were first mentioned in the submission instructions at NDSS in 2006, but this was also not widely adopted until 2017.

Abstract registration. Requiring authors to register the title and abstract of their submission in advance of the submission deadline is a practice that has recently reappeared. Figure 5 shows that most conferences experimented with this policy a decade ago, ultimately dropping it, but have reintroduced it in the past year or two.

Turnaround time. Based on the collected dates for different stages of the submissions, we quantify the number of days between paper submission and notification of acceptance in Figure 6, showing a gradual increase over time.

Short-lived experiments. While many policies remain in the CfPs after their introduction, some get removed after a few editions. Notable examples include S&P offering submission deadlines in every month between 2018 and 2020, and the practice of requiring authors to attach prior reviews to their (re)submissions, which was in place at S&P, USENIX SEC, and ACM CCS, but has been dropped by all of them.

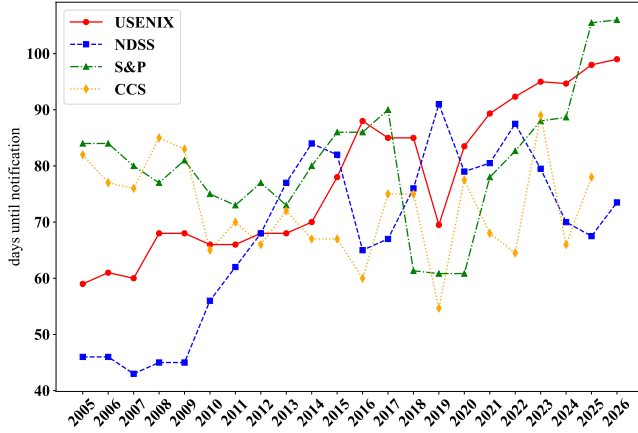


Figure 6. Days between paper submission and acceptance notification.

4. Future work and applications

4.1. Using our data

We believe that conference metadata, as explored in CyCoAnalysis, can provide the basis for rich meta-science research and informed policy decisions at future events.

In research, our dataset can be used to measure community growth, diversity, and inclusivity, exposing biases and the impact of policy decisions on shaping the community. Future work could examine whether the factors associated with publishing at conferences (publishing fees, conference tickets, conference locations, grants, and discounts offered) have an impact on author demographics for a given conference edition. Additional factors that can incur or eliminate biases in the set of authors are the enforcement of anonymous submissions and declaring conflicts of interest. The open access status can not only influence who decides to publish at a given venue, but it might also have an impact on the visibility of papers and the number of citations, as has been shown in other research fields [4]. The composition of PCs and conference organizers could also be studied more in depth, e.g., analyzing the overlaps between different PCs, how much experience (vice-)chairs gain before taking on this role, and whether there is a rolling system for program chairs to ensure knowledge transfer.

While topics of interest have been provided at most conference editions, this has been abandoned by CCS in 2008. By applying topic modeling to the published papers [16], the influence of the list of topics could be investigated. A deeper analysis of topics could also uncover which subfields of security have persisted over time and which topics have faded out over the years.

Ethics, open science, and AI usage are hot topics of discussion, with, e.g., recent editions of the USENIX SEC CFP devoting significant attention to these matters. Our dataset currently only classifies whether the CFP discusses these topics, but this could provide a starting point to a more in-depth analysis of how these policies impact the published papers, similar to Klemmer et al. [6].

For conference organizers specifically, we hope that our dataset can enable more informed and transparent decision-making, possibly even inspiring more community involvement, as it has recently been done for feedback on the presentation format of USENIX SEC 2025 [13] and the CFP of ACM CCS 2026 [8]. Conducting these initiatives with version control provides the added benefit of allowing easy tracking of changes from one conference edition to the next. We also encourage conference organizers to share more data with the community, such as the USENIX SEC 2026 chairs’ transparency report [14], elements of which could be integrated in CyCoAnalysis.

4.2. Extending our data

Our vision is to extend the coverage of the dataset (cf. Section 2) in all dimensions: adding more conferences, extending the coverage in time, and including further metadata across the board. In ongoing work, we are studying how artifact sharing in the literature is affected by conference policies and artifact evaluation processes. As part of this project, we will extend CyCoAnalysis with more details on the artifact evaluation policies at the studied conferences. Additionally, we are planning to include more detailed information on the outcomes of the review process (total submissions, early rejection, minor/major revisions), with corresponding statistics, which are often shared as part of the chairs’ message in the published proceedings. We would also like to refine how rebuttals are defined in the dataset, as historically this has covered different concepts: appealing an early rejection decision, responding to only the first round of reviews, or addressing all reviews. Finally, additional metadata of interest include paper formatting requirements, limits on the number of submitted papers, allowing the publication of pre-prints and other publicity during the review process, organizing and steering committees, associated workshops, and many more. For obtaining additional data, automatic tools could also be considered with sufficient human validation.

4.3. Community engagement

We would like CyCoAnalysis to become a community project, used and extended by independent researchers. To facilitate this community engagement, we propose a two-pronged approach. First, we are releasing a website, <https://cycoanalysis.github.io>, which makes the data easy to browse and download to be used in research. Second, we are investing in technical infrastructure to allow a smooth contribution process to the dataset. All our data is represented in YAML files with corresponding JSON schemas that define their structure. These are validated in CI, ensuring that pull requests conform to the format. At the same time, we encourage structural extensions to the dataset (e.g., adding a new metadata field to all conferences), which entails changing the schema to ensure future compatibility.

5. Conclusion

We presented CyCoAnalysis, a dataset about cybersecurity conferences. This project is envisioned as a long-term community effort, providing the basis for different branches of meta-science research in cybersecurity and aiding conference organizers in making informed policy decisions, indirectly benefiting all of the cybersecurity research community. CyCoAnalysis is available and open to contributions at <https://cycoanalysis.github.io>.

Acknowledgments

We thank Victor Le Pochat, Daan Vansteenhuyse, and the anonymous reviewers for their feedback on this paper. This research was partially funded by the Internal Funds KU Leuven and the Cybersecurity Research Program Flanders.

References

- [1] ACM, “Open access at ACM,” 2025. [Online]. Available: <https://authors.acm.org/open-access>
- [2] D. Balzarotti, “System security circus,” 2023. [Online]. Available: <https://www.s3.eurecom.fr/~balzarot/security-circus/>
- [3] A. Crowder, A. Lu, K. Childs, C. Stillman, P. Traynor, and K. R. Butler, “Data to infinity and beyond: Examining data sharing and reuse practices in the computer security community,” in *IEEE Symposium on Security and Privacy (S&P)*, May 2025, pp. 2678–2696. [Online]. Available: <https://ieeexplore.ieee.org/document/11023396/>
- [4] G. Eysenbach, “Citation advantage of open access articles,” *PLOS Biology*, vol. 4, no. 5, May 2006. [Online]. Available: <https://doi.org/10.1371/journal.pbio.0040157>
- [5] G. Gu, “Computer security conference ranking and statistic,” 2022. [Online]. Available: https://people.engr.tamu.edu/guofei/sec_conf_stat.htm
- [6] J. H. Klemmer, J. Schmäser, F. Fischer, J. Suray, J. Holtgrave, S. Lenau, B. M. Lowens, F. Schaub, and S. Fahl, “How transparent is usable privacy and security research? A meta-study on current research transparency practices,” in *USENIX Security Symposium, 2025*, pp. 5967–5986.
- [7] V. Le Pochat, “Reflecting on research practices,” *Communications of the ACM*, vol. 67, no. 5, pp. 37–39, May 2024. [Online]. Available: <https://doi.org/10.1145/3651965>
- [8] Z. Lin, “Request for feedback on the ACM CCS 2026 draft CfP (via GitHub issues),” 2025. [Online]. Available: <https://github.com/ACM-CCS-2026/CfP>
- [9] D. Olszewski, A. Lu, A. Crowder, N. Bennett, S. Layton, S. H. Varma Bhupathiraju, T. Tucker, S. Kalgutkar, H. Ver Helst, C. Stillman, K. R. B. Butler, S. Rampazzi, and P. Traynor, “Reproducibility in applied security conferences: An 11-year review on artifacts and evaluation committees,” in *ACM Conference on Reproducibility and Replicability (ACM REP)*, Jul. 2025, pp. 96–107. [Online]. Available: <https://dl.acm.org/doi/10.1145/3736731.3746151>
- [10] D. Olszewski, A. Lu, C. Stillman, K. Warren, C. Kitroser, A. Pascual, D. Ukirde, K. Butler, and P. Traynor, ““Get in researchers; we’re measuring reproducibility”: A reproducibility study of machine learning papers in tier 1 security conferences,” in *ACM Conference on Computer and Communications Security (CCS)*, 2023. [Online]. Available: <https://doi.org/10.1145/3576915.3623130>
- [11] D. Olszewski, T. Tucker, K. R. B. Butler, and P. Traynor, “Sok: Towards a unified approach to applied replicability for computer security,” in *USENIX Security Symposium, 2025*, pp. 469–488.
- [12] M. Payer, “An overview of the systems circus,” 2026. [Online]. Available: <https://nebelwelt.net/pubstats/>
- [13] G. Pellegrino, “Usenix security 2025 conference format - GitHub,” 2024. [Online]. Available: <https://github.com/USENIX-Security-2025/conference-format>
- [14] E. Redmiles and B. Stock, “Usenix security ’26 between-cycle transparency report,” 2026. [Online]. Available: https://docs.google.com/document/d/e/2PACX-1vSnHpeTFReelMZ124SzmGObXrHDN_B-UgXhOaE2RKfnfHn_qWuob17kPg4TKBMhhZy_qxYtggCo7vjK/pub
- [15] K. Rieck, “Influential security papers,” 2025. [Online]. Available: https://mlsec.org/topnotch/sec_papers.html
- [16] M. Schloegel, D. Klischies, S. Koch, D. Klein, L. Gerlach, M. Wessels, L. Trampert, M. Johns, M. Vanhoef, M. Schwarz, T. Holz, and J. Van Bulck, “Confusing value with enumeration: Studying the use of CVEs in academia,” in *USENIX Security Symposium*, Aug. 2025.
- [17] sec-deadlines, “Security and privacy conference deadlines,” 2026. [Online]. Available: <https://sec-deadlines.github.io/>
- [18] secartifacts, “Security research artifacts,” 2026. [Online]. Available: <https://secartifacts.github.io/>
- [19] A. Soneji, F. B. Kokulu, C. E. Rubio-Medrano, T. Bao, R. Wang, Y. Shoshitaishvili, and A. Doupé, ““Flawed, but like democracy we don’t have a better system”: The experts’ insights on the peer review process of evaluating security papers,” in *IEEE Symposium on Security and Privacy (S&P)*, 2022, pp. 1845–1862.
- [20] J. Zhou, “Top cyber security conferences ranking,” 2025. [Online]. Available: <http://jianying.space/conference-ranking.html>