Universität der Bundeswehr München

CISS MUNICH WORKING PAPER SERIES
WORKING PAPER 003/2025

CISS
Center for Intelligence & Security Studies

CYBER THREATS AND CRISIS EARLY WARNING

ISSN 3053-7800

Summary

Digital technologies and increasing global interdependence are transforming the nature of modern conflict. The Russian full-scale invasion of Ukraine has intensified attention on so-called hybrid threats—coordinated actions that combine conventional and unconventional means, including cyber attacks. State-targeted cyber operations, especially those aimed at critical infrastructure, are difficult to detect, cross national borders, and often escape traditional conflict monitoring. Reliable data and deep contextual understanding are essential for anticipating these threats. This study evaluates the forecasting potential of publicly available cyber incident datasets, identifying the European Repository of Cyber Incidents (EuRepoC) as the most suitable source for analyzing politically motivated cyber attacks against states. Drawing on similarities with conventional conflict data, the study outlines key hypotheses for cyber threat prediction. It proposes advanced forecasting methods such as spatio-temporal graph models and ensemble techniques. The overarching aim is to assess the feasibility of cyber-specific early warning systems and contribute to the development of predictive tools tailored to state-targeted cyber threats in an increasingly digital conflict landscape.

Stephanie Riedl, Lena Runge

Introduction

The timeliness and urgency of the discussion on anticipating, detecting and responding to cyber attacks in contemporary conflicts has been fueled by several broadly known incidents caused by both state and non-state actors. For the former, the 2010 Stuxnet cyber attack serves as one of the earliest examples when a cyber attack - which was suspected to be initiated by Israel and the United States of America - damaged centrifuges at an Iranian nuclear facility (Alvarez 2015). For the latter, more recent examples are the cyber operations conducted by the decentralized, international hacktivist group known as Anonymous against Russian actors in the context of the full-scale Russian invasion of Ukraine since 2022 (Tidy 2022). Generally, technological innovations in the field of digitalization, the information sphere, and the Internet of Things accelerated both progress and exposure to new vulnerabilities in today's conflicts. Consequently, the ability to probabilistically estimate the occurrence and frequency of state-targeted cyber attacks in complex, comprehensive conflict settings has become a pressing priority in the public sector to anticipate and mitigate threats in and from cyberspace. Whereas the prediction of physical conflict events gained increasing popularity and impetus in times of great computational advancements, less attention has been paid to other target variables like cyber events. Structural analysis of the apparent correlation of cyber and conflict in today's hybrid security and operational environment as well as synergies in event prediction is lacking. Therefore, this working paper aims to discuss the potential and pitfalls of using cyber data in forecasting the frequency of state-targeted, state-sponsored, or politically motivated cyber attacks. For this, this working paper explores the applicability of conflict forecasting methods to the cyber domain and highlights the predictive potential embedded in the temporal dependencies of cyber event data. In order to do so, this working paper analyses (1) selected publicly available cyber datasets to determine the data to work with, as well as (2) temporal dependencies in the cyber event data provided by the European Repository of Cyber Incidents (EuRepoC). The analysis also highlights (3) parallels between physical conflict data in the

Armed Conflict Location and Event Data (ACLED) and cyber attacks, adding to the immense scope or research on and with ACLED (Vesco et al. 2022; Trivedi et al. 2021; ACLED 2025a, 2025b; Raleigh et al. 2023). Based on the insights gained, promising directions for future research will be outlined to guide further empirical investigation. Based on data exploration, this study proposes a forecasting approach based on spatio-temporal models and graph neural networks. Through this, the working paper contributes to the emerging field of cyber event forecasting in several ways. First, it offers a systematic comparison of publicly available cyber event datasets, evaluating their quality, structure, and applicability for state-level early warning and conflict research. Second, it conceptually links insights from conflict forecasting — such as temporal dependencies and ensemble modeling — to the cyber domain. Third, it presents exploratory empirical findings that reveal temporal autocorrelation in state-targeted cyber attacks and correlations with physical conflict events. Finally, it identifies methodological challenges related to attribution, zero-inflation, and data harmonization, and outlines directions for future research to address these gaps.

Literature on Cyber Threats and Cyber Event Prediction

Literature on cyber in the context of today's conflicts revolves in some parts around understanding typologies of cyber events. According to Leigher (2021), referring to the gravity of intervention and the purpose, offensive cyber acts can be roughly categorized in three types - cyber espionage, cyber crime and cyber warfare (Leigher 2021, p. 8). Each type of cyber attack can appear in different forms. While data theft and data hijacking mostly fall under cyber espionage, system disruption for example via disruptive DDoS attacks (Mayne 2024), and malware that encrypts systems, files, or data for blackmail (ransomware) often exceeds mere espionage (EuRepoC 2025). Principally, cyber espionage and likewise cyber crime do not explicitly aim at destroying or disabling (virtual or physical) targets, whereas cyber warfare does to achieve political objectives (Leigher 2021, p. 8). In addition, literature on cyber deals with the diversity and number of actors involved and their motivations. Depending on the specific type of cyber operation, necessary expertise does not need to be highly sophisticated to have an impact. Consequently, respective acts can be executed by both state or non-state actors, by collective or individual actors, as well as by professionals as well as amateurs (Pawlak 2017, p. 2; Bewarder et al. 2025). Accordingly, the target group as well as actors for countering the threats and foster resilience vary, including state actors, civil society and individual citizens, as well as enterprises and the private sector (Pawlak 2017, p. 3; Tsetsos 2021, p. 9). In addition, other literature analyzes the threat potential of cyber in contemporary conflict. The nearly full integration of key areas in private and public life into cyberspace entails severe risks and vulnerabilities in society, economy, science, and state (Pawlak 2017, p. 8). For example, at the beginning of the full-scale Russian invasion of Ukraine the US-American cyber security company Cloudflare has shown how DDoS attacks accounting for over 80 percent of all ".ua"-domain traffic by early March 2022 could be related to the ongoing war (Cloudflare 2023).

Literature using cyber data most often decide for one data source without considering the potential of others (Devey 2024; Lakha et al. 2023). However, a recent study by Song et al. (2024) reviews and optimizes existing cyber datasets before using the CISSM Cyber Events Database to apply intervention analysis to detect significant shifts in cyber attack patterns and their correlation with major events. Through this, Song et al. find that the COVID-19 pandemic and the Russia-Ukraine war reshaped the cyber threat landscape - for the former on a global, for the latter on a regional scale (Song et al. 2024, p. 9, 13). Other literature on cyber data often deals with qualitative challenges of data availability, balance and completeness. In general, statistical analysis requires datasets to cover data in a constant manner over a long time capturing a significant number of incidents. In brief, the larger and the more informative a dataset is, the better for data-centric research approaches to analyze cyber events (Song et al. 2024). One of the reasons for data unavailability is that cyber attacks often concern security interests and consequently information under disclosure. Another reason is that often, steps to hide or obscure information impede the clear and unambiguous attribution of cyber incidents (Council on Foreign Relations 2025). So called cyber attack forensics requires a lot of time, financing, and a high degree of technical and investigative expertise (Boerbert 2010; Mueller et al. 2019). Hiding attribution of cyber attacks can also result from political factors. For instance, analyzing the EuRepoC data, Ella M. Devey (2024) showed that countries with a higher degree of press freedom are more likely to publicly address the attribution of a cyber attack to an actor (Devey 2024). Another reason for incompleteness and limited dataset quality are data or coding inconsistencies and respective error rates (Song et al. 2024, p. 4), as well as different reporting biases. For example, data on cyber is subject to language and identification biases and slightly over-represents English-speaking countries or countries with English media (Council on Foreign Relations 2025).

In view of the literature on cyber attack prediction, the forecasting of cyber attacks has received growing attention in the research community, driven by advances in machine learning and artificial intelligence. However, much of this work focuses on cyber attacks within a criminal framework, particularly targeting incidents directed at private companies (Ali et al. 2024). Companies share anonymized data on cyber incidents with researchers, offering a valuable foundation for quantitative analysis. The available data typically derives from network monitoring tools, web logs, or application logs, and are often characterized by high temporal granularity. Event counts for attempted cyber attacks can be recorded at the minute or even second level, enabling the application of a wide range of machine learning models, including deep learning techniques that require large volumes of data. Forecasting methods range from classical statistical time series models such as ARIMA and Markov Chains to more complex architectures like Long Short-Term Memory (LSTM) networks or other Deep Learning models (Landauer et al. 2025). Employing the idea of using public signals from online media, Okutan, Yang, and McConky (2018) demonstrate improved prediction performance for cyber attacks against private-sector companies by incorporating event mentions and sentiment metrics about global conflict from the GDELT database (GDELT 2022; Okutan et al. 2018). These features, which relate to companies and cyber-relevant technologies, serve as predictor variables and enhance model accuracy (Okutan et al. 2018).

In conflict research, several methodologies have been identified that are better suited to the characteristics of conflict event count prediction than other algorithms. The 2022 and 2023/2024 VIEWS (Violence Early Warning System) challenge aimed to forecast fatalities across the African continent - in 2023/2024 additionally for the Middle East - at both the country level and the standardized spatio-temporal global grid structure PRIO-GRID at a resolution of 0.5 x 0.5 decimal degrees (Tollefsen et al. 2012; PRIO 2016) - using machine learning models submitted by researchers worldwide. A summary of the lessons learned by Hegre, Vesco, and Colaresi (2022) highlights that models incorporating temporal dynamics — such as LSTMs or Markov Chains — performed best, alongside more complex models utilizing graph neural networks. Overall, ensemble forecasts combining multiple algorithms consistently yielded the most accurate results (Hegre et al. 2022, p. 547). At the PRIO-GRID level, ensembles of hurdle regression models — designed to account for the zero-inflated nature of conflict data — demonstrated the highest predictive performance (Mittermaier et al. 2024).

By contrast, relatively little research has addressed cyber attack forecasting at the geopolitical level in systematic quantitative manner, particularly with the aim of anticipating state-targeted attacks. A recent study by Roumani and Alraee (2025) goes beyond the qualitative description of cyber attacks by examining the severity of cyber attacks on critical infrastructure. Using data from the EuRepoC dataset, the authors apply ordinal regression to investigate the factors that influence the severity of the attack (Roumani and Alraee 2025). Their findings suggest that incidents involving ransomware tend to escalate the severity of cyber attacks (Roumani and Alraee 2025). They hypothesize that this is due to improved security protocols for other so-called Pertinent Alleviation Technologies (PATs), whereas ransomware remains highly lucrative for cybercriminal groups (Roumani and Alraee 2025, p. 8). With regard to attacker type, the study reveals that only state-led cyber attacks significantly increase the level of damage inflicted, underscoring the need to develop robust forecasting methodologies specifically tailored to state-sponsored threats (Roumani and Alraee 2025, p. 8). Also, yet to be fully explored are the underlying drivers of cyber attacks at the state level. As outlined before, Song et al. find in their recent analysis of the cyber landscape that major geopolitical and societal events had a significant impact on the dynamics of international cyber aggression (Song et al. 2024).

However, a few contributions have attempted to enhance cyber attack forecasting at the state level. In a study by Lakha et al. (2023), the authors identify periods of heightened risk for cyber attacks at the country, regional, and global level. They rely on datasets of state-level cyber incidents from the Cyber Operations Tracker and conflict-related event data from the Integrated Crisis Early Warning System (ICEWS) (Lakha et al. 2023; García-Durán et al. 2018). Their approach involves generating news-based political event graphs using the SIR-GN framework (Lakha et al. 2023; Layne and Serra 2021). These graph-based representations encode the structure of overall international conflict and they are integrated into a forecasting algorithm for anomaly detection. The study finds that both static and dynamic political event graphs derived from news sources can, in certain cases, enhance the predictive accuracy of cyber attack models (Lakha et al. 2023).

A further approach focusing on state-level cyber event forecasting is presented by Almahmoud et al. (2025). The authors employ custom web scraping tools to detect emerging trends in PATs, such as ransomware and DDoS attacks, as well as developments in state-targeted cyber incidents based on information from government reports, news media, and social platforms. These trends are then used as inputs for Bayesian spatio-temporal Graph Neural Networks (Scarselli et al. 2008) - as first proposed by Wu et al. (2020) - which aim to predict cyber threat patterns based on the evolution of relevant technologies (Almahmoud et al. 2025). However, a statement regarding the completeness of their dataset on cyber attacks compared to sources as EuRepoC or the CISSM Events Database is lacking. Therefore, the authors might miss cyber incidents that were obtained in specialized web-scraping algorithms for cyber incidents like the one used in EuRepoC. In total, research increasingly revolves around cyber attacks and conflict analysis. Its growing relevance is stressed by the availability of different open-source datasets on cyber incidents with a political dimension. While the quality of datasets vary, they offer a valuable starting point for further analysis. Besides, existing forecasting research predominantly targets private-sector cyber incidents or general threat trends with limited attention to systematic, state-level prediction. No comprehensive modeling framework currently exists for anticipating state-targeted cyber attacks using quantitative methods, highlighting a significant gap in literature and a need for further empirical development.

Empirical Exploration of Cyber Event Data and Prediction Approaches

In the context of a first empirical exploration of cyber event data and prediction of cyber events, the following section compares publicly available datasets on cyber events and examines empirical potentials of forecasting cyber events.

Datasets on Cyber Events

For the development of robust models to predict cyber attacks data on cyber is indispensable. Event data on cyber appears mainly in two contexts - (1) either in specialized datasets that focus on cyber incidents entirely, or (2) in unspecialized datasets in which cyber is a subtype among many other types of conflict-related events. Regarding the latter two prominent examples are the Global Database of Events, Language, and Tone (GDELT) project and Political Event Classification, Attributes, and Types (POLECAT). The GDELT Event Database codes event data on cyber events, for example the mobilization or increase of cyberforces (CAMEO 155) or a cybernetical attack (CAMEO 176) (Schrodt 2012, p. 137). An example on how GDELT is used in the analysis of cyber events is provided by Anna Piazza and Srinidhi Vasudevan (2025) who investigate the development of cyber threats in healthcare as a critical infrastructure (Piazza and Vasudevan 2025). Likewise, POLECAT provides context data on discussions of cyber, including cyber attacks, cyber security, cyber warfare, and cyber crime (Halterman et al. 2023a, p. 15). For example, Halterman et al. (2023) analyzed the distribution of context labels in PLOVER in the time between June 2020 and February 2023. Of a total of 59.2 percent events with context, 0.8 percent included cyber (Halterman et al. 2023b, p. 22). However, as the cyber data from the conflict event datasets need extensive data pre-processing, this working paper only focuses on cyber event datasets in detail comparing the European Repository of Cyber Incidents (EuRepoC), the Cyber Operations Tracker provided by the Council on Foreign Relations, and the Center for International and Security Studies at Maryland (CISSM) Cyber Events Database.

Overall, there has been few research examining the strengths and weaknesses of different public cyber databases for statistical analysis and other data-centric approaches. Principally, not all cyber acts warrant inclusion in state-level security analysis. For that, a specific focus lies on politically motivated and state-sponsored cyber attacks. Moreover, sophisticated quantitative research - like event forecasts - requires sufficient data of high quality. Comparing the three cyber event datasets outlined above, Table 1 provides a brief overview over some similarities and differences.

Category	EuRepoC	Cyber Operations Tracker	CISSM
Temporal Coverage	Since 2000	Since 2005	Since 2014
Spatial Coverage	worldwide	worldwide	worldwide
Event Focus	political	State-based	Political, economic, other
Total Events (05/25)	3414	917	14557
Categories	84	11	16
Examples	Receiver, initiator, types, conflict issue, attribution, responses	Types, sponsor, actor's affiliation, victim, response	Type, actor, motives, target, attribution
Updates	continuously	quarterly	monthly

Table 1: Public Cyber Datasets in Comparison

The three datasets vary quite considerably. The differences make a standardized encoding across datasets and data integration rather difficult, raising the question about the added value of such a processing step. To illustrate, all three datasets provide information about the initiating or attacking actor(s) but in different manners. The Cyber Operations Tracker includes the initiating actor itself (”affiliations”, e.g. Lazarus Group) and the actor’s sponsoring state (”sponsor”, e.g. Korea (Democratic People’s Republic of)). This category implies the state-based or state-affiliated nature of the events. Compared to that, the CISSM Cyber Event Database includes the initiating actor itself (”actor”, e.g. Syrian Electronic Army), its type (”actor type”, e.g. criminal), and the actor’s country (”actor country”, e.g. Syria). Like this, the country refrains from automatically implying an actor’s sponsorship. Likewise, EuRepoC includes the initiating actor’s name (”initiator name”, e.g. Liazor), the country it is from (”initiator country”, e.g. Armenia), its group type (”initiator category”, e.g. non-state-group), and its group subtype (”initiator subcategory”, e.g. hacktivist(s)). Moreover, regarding the attribution, there are significant differences. While the CISSM Cyber Events Database only copies the attribution from the source without any additional validation, EuRepoC’s process involves their researchers to manually code the data. Furthermore, there seems to be coding inconsistencies due to arbitrariness in the data. To illustrate, although both record the same DDoS attack on Ukrainian government websites in February 2021 the CISSM Cyber Event Database does not assign an actor whereas EuRepoC assigns Russian threat actors to be the responsible actor as shown in Table 2. For a cyber attack on the Indian power grid infrastructure in November 2020, CISSM Cyber Event Database clearly states RedEcho from China the responsible actor while EuRepoC categorizes the actor to be unknown. At the same time, the attack on the US-American agency websites in December 2019 have been assigned to the same actor, being Iranian cyber units.

Date	Cyber Attack On	Actor - EuRepoC	Actor - CISSM
Feb 2021	UKR gov. websites	RUS threat actors (RUS)	Undetermined (Unknown)
Nov 2020	IND power grid infra.	Unknown (Undetermined)	RedEcho (CHN)
Dec 2019	USA agency websites	IRN cyber units (RN)	IRN cyber units (IRN)

Table 2: Cyber Events in Both EuRepoC and the CISSM Event Database

Besides, initial attempts to match cyber incidents among all three datasets based solely on the reported start date yielded no matches. These findings suggest a more sophisticated event alignment approach, potentially leveraging large language models (LLMs) to compare detailed incident narratives, threat actor profiles, and operational characteristics across datasets as well as to enhance dataset interoperability and enable more robust cross-validation of cyber incident records.

At this point a dataset integration does not seem feasible and a choice between the datasets is inevitable to proceed with the analysis of cyber and conflict. On a conceptual level, due to its comprehensiveness EuRepoC seems to excel among the datasets. Conceptually, it covers the longest time, offers most frequent updates and provides most categories. For example, EuRepoC is the only dataset that intends to link cyber events to offline conflicts or to add information on the incident’s impact. Moreover, EuRepoC is the only dataset that includes detailed information on the reason for the integration of an incident (inclusion criterion) and for the attribution. Comparing the response categories, although the Cyber Operations Tracker focuses on a limited selection of six political responses, EuRepoC offers a broad variety of legal and political responses. EuRepoC’s focus on the political dimension - including politicized or non-politicized attacks on political targets, politicized attacks on non-political targets, and attacks on critical infrastructures conducted by either a state or state-affiliated actor or politically motivated non-state actor (EuRepoC 2023, p. 1) - is useful when addressing political dimensions of cyber events even if the total number of recorded events is only a fraction in contrast to the CISSM Cyber Events Database and thus struggles to meet statistical preference for large datasets.

Apart from the conceptual level, all three datasets suffer from different amounts of missing, unknown or unavailable data affecting data quality. As relevant information is often difficult to obtain, dataset concepts regularly diverge from their feasible implementation. This imposes challenges for potential future analysis when using their indicators and labels. For EuRepoC, this affects among others data on responses, impact, attribution as well as concrete names of the targeted entities. Likewise, the Cyber Operations Tracker underlines uncertainties, for instance marking only possible or suspected links in actor's affiliation. In contrast to challenges of physical conflict event data in quantitative conflict research, data on cyber events pose additional difficulties due to the numerous actors, attribution issues, irregular patterns and the rapidity of cyber incidents. Based on the selection of EuRepoC as the primary data source, several empirical results regarding the time series behavior of cyber attacks will be presented in the following section as well as implications for suitable forecasting approaches.

Analyzing & Forecasting State-targeted Cyber Attacks

For the analysis, all EuRepoC entries attributed to at least one state on a given day are considered. In cases where a cyber incident is attributed to multiple states, an event count of one is assigned to each attributed state. To explore the relationship between cyber and physical conflict, EuRepoC is combined with ACLED by matching events based on country code and event date.

Similar to political violence data in ACLED - where daily events are categorized into types such as battles, violence against civilians, protests, or explosions (ACLED 2025d) - cyber attack incidents display rare-event behavior, posing significant challenges for predictive modeling (Mittermaier et al. 2024). The highly sparse and non-normally distributed nature of these events reduces the effectiveness of many traditional machine learning techniques (Hegre et al. 2022). Table 3 presents a summary of daily event counts across countries, comparing political conflict events in ACLED with cyber incidents in EuRepoC.

Dataset	ACLED (Conflict)	EuRepoC (Cyber)
Data Coverage	2020-01-01 to 2025-05-07	2000-01-01 to 2025-05-05
Zero-Inflation	61.039	99.756
Mean	3.411	0.003
Median	0.0	0.0
Std Dev	12.241	0.079
Min	0.0	0.0
10%	0.0	0.0
90%	8.0	0.0
Max	924.0	22.0
Skewness	10.811	72.717
Kutosis	249.621	11319.736
Country Coverage	210 countries and regions	169 countries

Table 3: Summary Statistics for Daily Event Counts (ACLED and EuRepoC) Across All Avalable Countries and Regions

The comparison highlights that both datasets are highly zero-inflated. While 61 percent of all country-day observations in ACLED are associated with zero events, the proportion rises to 99.76 percent for cyber events in EuRepoC. This extreme sparsity poses an additional challenge for developing reliable cyber forecasting models. Additionally, ACLED documents conflict events in 210 countries and territories, including de facto states such as Palestine and Kosovo (ACLED 2025c), whereas cyber incidents in EuRepoC are recorded for only 169 countries.

Despite these challenges, promising statistical properties can be identified. A time series examination of the EuRepoC dataset reveals strong temporal autocorrelation: the number of incidents in any given week is substantially influenced by the number of incidents in the preceding week. Autocorrelation means past cyber activity increases the likelihood of future attacks. This is useful for building alert systems that can detect threat escalation patterns. Figure 1 illustrates that, on average, the weekly number of detrended cyber incidents targeting the Ukrainian state was correlated at approximately 40 percent with the number of incidents in the preceding week — exceeding the significance threshold. This pattern remains robust across the top ten countries most-affected by cyber attacks, even after accounting for long-term upward trends through differencing. Differencing is a common method in time series analysis used to remove long-term trends. This allows for a clearer assessment of whether short-term patterns—such as week-to-week correlations—still exist once the general increase in cyber attacks over time is accounted for. These findings indicate that, despite sparsity, meaningful temporal structures exist in the data that could be leveraged for predictive modeling. This finding aligns with conclusions from conflict research, which emphasize that past conflict remains one of the strongest predictors of future violence (Hegre et al. 2022).

In addition to temporal dependencies, cyber incidents also appear to exhibit spatial correlation across countries, suggesting the presence of inter-state effects and relational dynamics. However, such correlations must be interpreted with caution, as they represent only a first step toward causal inference. Observed patterns may partially result from the structure of cyber operations themselves, for example attacks launched by the same attackers are often coordinated to target multiple international organizations simultaneously, resulting in multi-country attribution. Similarly, when institutions such as NATO are affected, attribution may be distributed across several member states, further complicating efforts to assign responsibility or analyze country-level targeting patterns. Whether spatial correlations persist when controlling for simultaneous multi-target attacks has to be confirmed by further analysis.

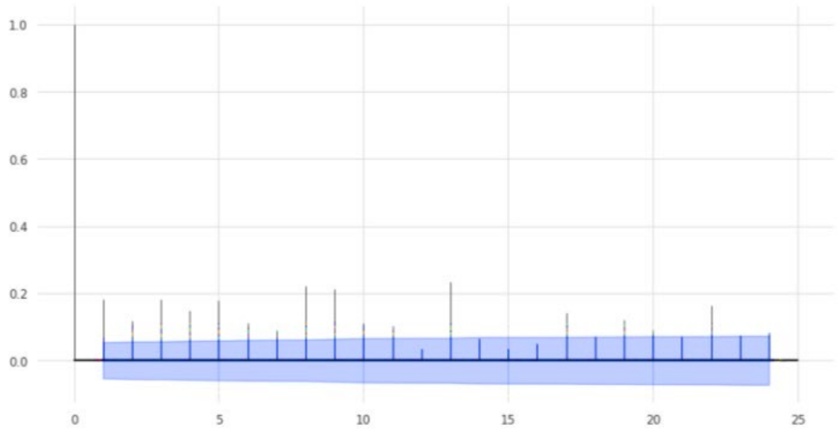

Figure 1: Weekly Autocorrelation of Cyber Attacks targeting the Ukrainian State. Y-scale shows magnitude while blue area indicates significance level. X-scale shows the weekly lag.

Nevertheless, these initial findings offer important starting points for future research on the spatial dynamics of cyber aggression. Furthermore, the findings suggest a tangible link between cyber attacks and physical conflict at the global level. When combining EuRepoC data with ACLED conflict data by matching on ISO3 country codes and event dates, a correlation is observed between the weekly frequency of cyber attacks and concurrent episodes of political violence. As shown in Figure 2, spikes in cyber incidents targeting the Ukrainian state coincide with the onset of physical conflict in early 2022. While this relationship requires further empirical investigation, it supports the notion that cyber operations often accompany — or function as extensions of — physical conflict, enabling states to exert pressure and cause disruption through cyber means. Building on the approach by Almahmoud et al. (2025), where Bayesian spatio-temporal graph neural networks (GNNs) were employed to forecast cyber threats using counts of online mentions of specific attack types, a similar methodology could be adapted to forecast EuRepoC cyber attack counts. These methods are able to learn inter-series correlations and temporal dependencies in the graph (Wu et al. 2020). This would involve constructing a graph based on historic cyber incidents and conflict events to model the interplay between state-targeted cyber attacks and physical conflict dynamics. As shown in the previous literature on cyber attacks (focusing on private companies), information on ongoing (physical) conflict in the target country can further improve predictive performance of cyber attack forecasting models. Such graph-based models could be benchmarked against more easily implementable ensemble approaches, such as XGBoost (Extreme Gradient Boosting) - which is an efficient and scalable machine learning algorithm based on gradient-boosted decision trees (Chen and Guestrin 2016) - extended with zero-inflated hurdle model variants to account for the high proportion of zero-event days in the EuRepoC dataset.

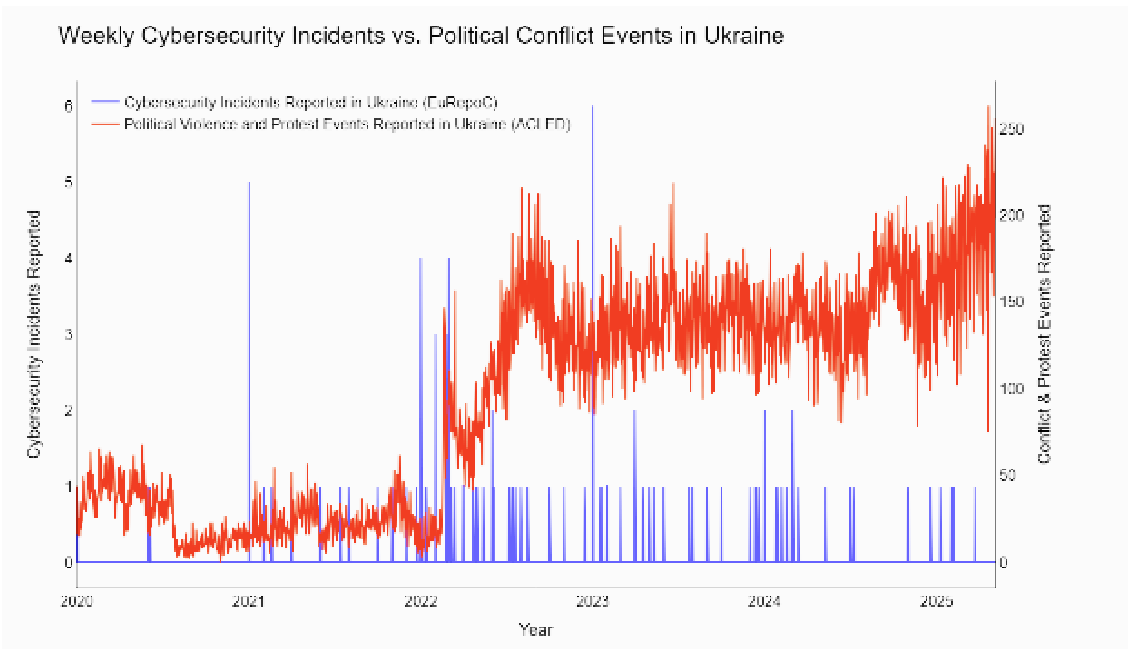

Figure 2: Weekly Cybersecurity Incidents in Ukraine (EuRepoC) vs. Political Conflict Events in Ukraine (ACLED).

Implications and Conclusion

This working paper sets out to describe and compare cyber datasets, evaluate their predictive potential for crisis early warning, and hence propose a methodological framework for the prediction of state-targeted cyber attacks. This was achieved through dataset comparison, exploratory analysis of EuRepoC - identified as the most promising dataset for forecasting state-targeted cyber attacks - and the transfer of conflict forecasting techniques to the cyber domain. Similar to physical conflict data, the analysis of cyber data reveals a high degree of zero-inflation as well as temporal dependencies in cyber attacks, expressed through autocorrelation. Drawing on insights from the literature, the use of graph-based spatio-temporal models is recommended for cyber attacks prediction. In addition, incorporating information on ongoing physical conflict and attack severity to the models may help to enhance predictive performance.

In summary, comparable data structures and statistical distributions - including strong temporal autocorrelation and potential spatial dependencies - between cyber attacks and physical conflict are observable and suggest notable parallels. Accordingly, methodological insights from traditional conflict prediction — such as the use of ensemble regression models — may be effectively applied to the forecasting of weekly or monthly state-targeted cyber attacks. However, particular caution must be exercised due to the highly zero-inflated nature of cyber event data, which poses even greater challenges for prediction than those encountered in the context of physical conflict. The empirical work on the prediction of cyber attacks is also bound to a couple of limitations: The often multinational nature of cyber attacks introduces attributional uncertainty. A single attack may be attributed to multiple states or organizations, resulting in highly correlated records. Methodological frameworks must account for redundancy and over-reporting, as same attacks might be targeted at different states or institutions. Furthermore, targeting might be ambiguous or incorrectly labeled. In this context, graph-based models may offer an advantage, as they are capable of representing complex actor relationships beyond state-based tabular structures. Given the low frequency of cyber attacks for most states, particular care must be taken to avoid overfitting and to validate models across diverse temporal and spatial contexts. Lastly, unlike physical conflict, where fatalities provide a clear outcome metric, cyber incident impact is difficult to quantify. For example, EuRepoC intends to provide data on different types of impact and intensity but in fact, data is often unavailable. At the same time, the number of (attempted) attacks alone does not necessarily reflect the threat level. As such, models should incorporate severity indicators or alternative proxies for impact.

Despite growing attention to cyber threats in the research community, state-level cyber attack forecasting remains underdeveloped. A systematic review of appropriate methodologies, datasets, and target definitions is needed to advance this field. However, in the cyber domain, these entities are often difficult to delineate, as cyber actors operate across borders and are not confined by conventional geographic boundaries. This might put the usefulness of some concepts based on physical conflict into question for investigating and predicting cyber events. Despite that and as the availability and quality of relevant data continue to improve, there is growing potential to better understand the role of cyber attacks within the framework of crisis early warning systems. For example, public data sources such as Cloudflare Radar (Cloudflare 2025) offer additional potential for enhancing early warning capabilities by monitoring real-time shifts in internet traffic linked to cyber threats (Cloudflare 2025). This may ultimately shed light on their interaction with, and possible connection to, ongoing physical conflicts and broader geopolitical dynamics.

Overall, by laying out the current limitations in cyber data, demonstrating methodological parallels to classic conflict forecasting, and identifying new research directions, it offers a foundation for future empirical work. To further advance the field, we encourage the expansion of hybrid threat datasets and the development of tailored prediction models that account for the unique dynamics of cyber operations in modern conflict environments. Developing predictive capacities in this field will not only advance academic knowledge but may eventually also contribute to more timely and effective responses to cyber-enabled crises in an increasingly interconnected world.

Stephanie Riedl, M. Sc., is a research associate at the Center for Crisis Early Warning at the University of the Bundeswehr in Munich (Germany). She has a background in statistics and computational social sciences. Her research focuses on time series prediction, conflict analysis and cyber threats.

Lena Runge is a research associate at the Center for Crisis Early Warning at the University of the Bundeswehr in Munich (Germany). She has a background in peace and conflict studies. Her research focuses on conflict analysis, international peace operations, and mixed methods.

References

ACLED (Ed.) (2025a): ACLED CAST: Conflict Alert System. Available online at https://acleddata.com/conflict-alert-system/, checked on 6/25/2025.

ACLED (Ed.) (2025b): ACLED: Bringing Clarity to Crisis. Available online at https://acleddata.com/, checked on 6/25/2025.

ACLED (Ed.) (2025c): Country and Time Period Coverage. Armed Conflict Location & Event Data Project (ACLED). Available online at https://acleddata.com/knowledge-base/country-time-period-coverage/, checked on 6/26/2025.

ACLED (Ed.) (2025d): General Methodology. Available online at https://acleddata.com/article-categories/general-methodology/, checked on 6/25/2025.

Ali, D. B.; Belaoued, M.; Dawaliby, S. (2024): Anticipating Cyber Threats: Deep Learning Approaches for DDoS Attacks Forecasting. In : 2024 8th Cyber Security in Networking Conference (CSNet). Paris, France: IEEE, pp. 128–132.

Almahmoud, Zaid; Yoo, Paul D.; Damiani, Ernesto; Choo, Kim-Kwang Raymond; Yeun, Chan Yeob (2025): Forecasting Cyber Threats and Pertinent Mitigation Technologies. In *Technological Forecasting and Social Change* 210, p. 123836. DOI: 10.1016/j.techfore.2024.123836.

Alvarez, Joshua (2015): Stuxnet: The world's first cyber weapon. Edited by Standford Center for International Security, Cooperation. Available online at https://cisac.fsi.stanford.edu/news/stuxnet, checked on 5/2/2025.

Bewarder, Manuel; Flade, Florian; Pittelkow, Sebastian (2025): Russischer Geheimdienst soll hinter Brandsätzen in Luftfracht stecken. Edited by tagesschau. Available online at https://www.tagesschau.de/investigativ/ndr-wdr/russland-sabotage-106.html, checked on 5/2/2025 % This file was created with Citavi 6.18.0.1.

Boerbert, W. Earl (2010): A Survey of Challenges in Attribution. In : Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy. Washington, D.C.: National Academies Press.

Chen, Tianqi; Guestrin, Carlos (2016): XGBoost: A scalable tree boosting system. In : Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 785–794.

Cloudflare (Ed.) (2023): One year of war in Ukraine. Available online at https://blog.cloudflare.com/one-year-of-war-in-ukraine/, checked on 6/8/2025.

Cloudflare (2025): Cloudflare Radar.

Council on Foreign Relations (Ed.) (2025): Tracking State-Sponsored Cyberattacks Around the World: Methodology. Available online at https://www.cfr.org/cyber-operations/#OurMethodology, checked on 5/12/2025.

Devey, Ella M. (2024): Public Cyberattack Attribution and Domestic Political Considerations: An Analysis of State Decision Making. Master's thesis. Utah State University, Logan, UT. Available online at https://digitalcommons.usu.edu/etd/.

Eck, Kristine (2012): In Data We Trust? A Comparison of UCDP GED and ACLED Conflict Events Datasets. In *Cooperation and Conflict* 47 (1), pp. 124–141. DOI: 10.1177/0010836711434463.

EuRepoC (Ed.) (2023): Codebook. European Repository of Cyber Incidents (EuRepoC). Available online at https://eurepoc.eu/wp-content/uploads/2023/07/EuRepoC_Codebook_1_2.pdf, checked on 5/12/2025.

EuRepoC (Ed.) (2025): Dashboard. European Repository of Cyber Incidents (EuRepoC). Available online at https://eurepoc.eu/dashboard/, checked on 5/2/2025.

García-Durán, Alberto; Dumancic, Sebastijan; Niepert, Mathias (2018): ICEWS. Available online at https://paperswithcode.com/dataset/icews, checked on 6/25/2025.

GDELT (2022): The GDELT Project. Available online at https://www.gdeltproject.org/, checked on 5/13/2025.

Halterman, Andrew; Bagozzi, Benjamin; Beger, Andreas; Schrodt, Philip A.; Scraborough, Grace (2023a): PLOVER and POLECAT: A New Political Event Ontology and Dataset. Edited by SocArXiv Papers. Available online at https://osf.io/preprints/socarxiv/rm5dw_v1, checked on 5/14/2025.

Halterman, Andrew; Bagozzi, Benjamin; Beger, Andreas; Schrodt, Philip A.; Scraborough, Grace (2023b): Supplemental Appendix for PLOVER and POLECAT: A New Political Event Ontology and Dataset. Available online at https://osf.io/z6g39.

Hegre, Håvard; Vesco, Paola; Colaresi, Michael (2022): Lessons from an Escalation Prediction Competition. In *Conflict Management and Peace Science*, pp. 521–554. DOI: 10.1080/03050629.2022.2070745.

Lakha, Bishal; Duran, Jason; Serra, Edoardo; Spezzano, Francesca (2023): Prediction of Future Nation-initiated Cyberattacks from News-based Political Events Graph. In : Proceedings of the 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA): IEEE, pp. 461–470.

Landauer, Max; Skopik, Florian; Stojanović, Branka; Flatscher, Andreas; Ullrich, Torsten (2025): A Review of Time-Series Analysis for Cyber Security Analytics: From Intrusion Detection to Attack Prediction. In *International Journal of Information Security* 24 (3). DOI: 10.1007/s10207-024-00921-0.

Layne, Janet; Serra, Edoardo (2021): Inferential SIR-GN: Scalable Graph Representation Learning. In *arXiv preprint arXiv:2111.04826*. DOI: 10.48550/arXiv.2111.04826.

Leigher, William E. (2021): Cyber conflict in a hybrid threat environment: Death by a thousand cuts. Edited by Hybrid CoE (Hybrid CoE Paper, 10). Available online at https://www.hybridcoe.fi/wp-content/uploads/2021/11/Hybrid_CoE_Paper_10_Cyber_conflict_in_a_hybrid_threat_environment_WEB.pdf.

Mayne, Mitch (2024): Jahresrückblick 2024: Aktuelle Erkenntnisse und Aussichten für die Zukunft. Edited by Akamai. Available online at https://www.akamai.com/de/blog/security-research/2024-december-today-insights-tomorrow-outlook-2025, checked on 5/2/2025.

Mittermaier, Daniel; Bohne, Tobias; Hofer, Martin (2024): Forests of UncertainT(r)ees: Using Tree-based Ensembles to Estimate Probability Distributions of Future Conflict. Technical Report. Center for Crisis Early Warning, University of the Bundeswehr Munich.

Mueller, Milton; Grindal, Karl; Kuerbis, Brenden; Badiei, Farzaneh (2019): Cyber Attribution: Can a New Institution Achieve Transnational Credibility? In *The Cyber Defense Review* 4 (1), pp. 107–122. Available online at https://www.jstor.org/stable/26623070.

Okutan, Ahmet; Yang, Shanchieh Jay; McConky, Katie (2018): Forecasting Cyber Attacks with Imbalanced Data Sets and Different Time Granularities. In *arXiv preprint arXiv:1803.09560*.

Pawlak, Patryk (2017): Countering hybrid threats: EU-NATO cooperation. Edited by EPRS (Briefing). Available online at https://www.europarl.europa.eu/RegData/etudes/BRIE/2017/599315/EPRS_BRI(2017)599315_EN.pdf.

Piazza, Anna; Vasudevan, Srinidhi (2025): Mapping the Cyberthreat Landscape in Healthcare Using GDELT: A Multimethod Approach. In *Health security*. DOI: 10.1089/hs.2024.0016.

PRIO (2016): PRIO-GRID. Available online at https://www.prio.org/data/9, checked on 6/25/2025.

Raleigh, Clionadh; Kishi, Roudabeh; Linke, Andrew (2023): Correction: Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. In *Humanit Soc Sci Commun* 10 (1), p. 1. DOI: 10.1057/s41599-023-01626-w.

Roumani, Yaman; Alraee, Mais (2025): Examining the Factors That Impact the Severity of Cyberattacks on Critical Infrastructures. In *Computers & Security* 148, p. 104074. DOI: 10.1016/j.cose.2024.104074.

Scarselli, Franco; Gori, Marco; Tsoi, Ah Chung; Hagenbuchner, Markus; Monfardini, Gabriele (2008): The graph neural network model. In *IEEE Transactions on Neural Networks* 20 (1), pp. 61–80.

Schrodt, Philip A. (2012): CAMEO: Conflict and Mediatiion Event Observations: Event and Actor Codebook. Edited by Event Data Project. Available online at http://data.gdeltproject.org/documentation/CAMEO.Manual.1.1b3.pdf, checked on 5/13/2025.

Song, Uihyeon; Hur, Gimin; Lee, Sangjin; Park, Jungheum (2024): Unraveling the Dynamics of the Cyber Threat Landscape: Major Shifts Examined Through the Recent Societal Events. In *Sustainable Cities and Society* 103, p. 105265. DOI: 10.1016/j.scs.2024.105265.

Tidy, Joe (2022): Anonymous: How hackers are trying to undermine Putin. In *BBC News*. Available online at https://www.bbc.com/news/technology-60784526, checked on 6/2/2025.

Tollefsen, Andreas F.; Strand, Havard; Buhaug, Halvard (2012): PRIO-GRID: A Unified Spatial Data Structure - Peace Research Institute Oslo (PRIO). In *Journal of Peace Research* 49 (2), pp. 363–374. Available online at https://www.prio.org/publications/5110, checked on 6/9/2025.

Trivedi, Anusua; Keator, Kate; Scholtens, Michael; Haigood, Brandon; Dodhia, Rahul; Ferres, Juan Lavista et al. (2021): How to Handle Armed Conflict Data in a Real-World Scenario? In *Philosophy & Technology* 34 (1), pp. 111–123. DOI: 10.1007/s13347-020-00424-5.

Tsetsos, Konstantinos (2021): Neue hybride Bedrohungen. Edited by METIS (26). Available online at https://metis.unibw.de/assets/pdf/metis-studie26-2021_07-hybride_bedrohungen.pdf.

Vesco, Paola; Hegre, Håvard; Colaresi, Michael; Jansen, Remco; Lo, Adeline; Reisch, Gregor; Weidmann, Nils (2022): United they stand: Findings from an escalation prediction competition. In : International Interactions, 48 (4), pp. 860–896. Available online at https://www.prio.org/publications/13337, checked on 12/16/2024.

Wu, Zonghan; Pan, Shirui; Long, Guodong; Jiang, Jing; Chang, Xiaojun; Zhang, Chengqi (2020): Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks. In : Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA: ACM, pp. 753–763.