Research Statement

My research offers timely empirical measurements to explore cyberspace and its societal impact at scale, focusing on underground subcultures fostering online crime and harms. The resulting insights help us better understand online threats and inform policy decisions for online safety and security. My primary approach is data-driven, with research questions addressed by rigorous quantitative and qualitative measurements of large-scale real-world evidence.

I. Measuring Security and Online Wickedness

Externalities, such as Covid-19, may cause significant shifts and enduring changes in human behaviour, both offline and online. My recent work explores the effects of major incidents that *intensify* or *disrupt* online crime and harms.

▸ Intensifying Online Wickedness: The Pandemic and Armed Conflicts

Turning Up the Dial @ IMC'20 [1] evidences the significant influence of Covid-19 on illicit trading activity on the largest underground hacking forum. This empirical observation can be attributed to the increasing time spent online by individuals during lockdowns. The paper reveals that various forum users overcame the 'cold-start problem' – where new traders face difficulty settling transactions due to a lack of reputation, yet they cannot gain reputation without trading – by engaging in low-value exchanges to build their trustworthiness before developing larger trading volumes.

Getting Bored of Cyberwar @ WWW'24 [2] explores the involvement of volunteer hacktivists and low-level cybercrime actors in the Russia-Ukraine conflict. Although they promptly participated in targeting digital assets of both countries after the invasion using DDoS and defacement attacks, this intensification was short-lived, with a clear loss of interest after a few weeks. Their activity may cause immediately noticeable effects, but the impact was mainly propaganda dissemination instead of contributing to the 'hard' digital frontline. While popular narratives tend to overhype and conflate these actors with persistent state-sponsored hacktivists, we believe they should be considered separately.

Yet Another Diminishing Spark @ Under Review [3] compares defacement attacks seen in the Russia-Ukraine conflict to those in the Israel-Hamas war, discovering similar patterns peaking shortly after the war started. While attacks were two-sided in the case of Russia-Ukraine, they have been mostly one-sided in the Israel-Hamas war: most targeted Israel while no significant waves have hit Palestine, presumably as Palestine has far fewer sites, many of which are hosted overseas. The scale of attacks on Israel and Palestine has been much less than those on Russia or Ukraine.

▸ Disrupting Online Wickedness: Industry and Police Interventions

No Easy Way Out @ S&P'24 [4] examines a concerted effort to dismantle Kiwi Farms, the largest forum for online hate and harassment. We show that solely relying on deplatforming, even by swift actions of several competent tech firms, can be insufficient. The forum traffic and activity were quickly disrupted, but gradually recovered after a few months. Many users temporarily decamped to Telegram, but returned when the forum was back and became even more connected. The industry often does better than government actions, but this extraordinary event suggests that shutting down a dispersed community is unlikely to be effective if the censor cannot incapacitate or deter the key operators.

Assessing the Aftermath @ Under Review [5] studies a global takedown of DDoS-for-hire services (or *booters*) involving the FBI, the NCA, and the Dutch Police. The first wave on 14 December 2022 seized 49 domains, and 13 more were seized in the second wave on 5 May 2023. These domains were redirected to a police-deployed page hosted by us to collect access flows across booters. We found that 26 first-wave seized domains quickly returned under new domains, while all 13 second-wave seized ones reappeared. These emergence, however, failed to recover their visit traffic, with over 80% being lost. The global DDoS attack volume quickly declined by half, with the effect lasting for only six weeks.

▸ Data Licensing: Reproducibility and Extensibility

Reproducibility and extensibility are crucial to me. All data used in my work can be shared with researchers, enabling them to immediately pursue ideas to address real-world problems without spending months or years collecting data.

ExtremeBB @ ACL WOAH'23 [6] is a dataset for large-scale research into online hate, harassment, and extremism. Outside of Cambridge, it has been licensed to 67 scholars in 27 groups across 21 institutions in 7 countries. I co-authored PostCog @ EuroS&P WACCO'22 [7], an interactive 'search engine' enabling researchers, especial non-technical scientists, to analyse our data visually and straightforwardly. I contributed to a chapter outlining how we ethically collect and share cybercrime and extremist datasets, as part of the book Open Source Investigations in the Age of Google [8].

II. Cyber Attacks and Defenses

I believe the best way to safeguard things is to figure out how to attack them. I am particularly fascinated by fundamental research uncovering novel practical attack vectors, and research that develops tools to counter such threats.

Erebus @ S&P'20 [9] is a novel attack capable of partitioning the Bitcoin P2P network in a stealthy manner, without the victims realising they are being isolated. It leverages the advantages of big Internet entities such as ASes and ISPs to intercept and monitor thousands of 'shadow' malicious Bitcoin nodes. These can subsequently establish connections with legitimate nodes, gradually populating a large number of IP addresses into the victims' peering tables, and ultimately isolating them from the rest of the network. This vector also affects many other Bitcoin-based cryptocurrencies, such as Litecoin, Bitcoin Cash, and Dogecoin, necessitating a few protocol tweaks in the Bitcoin codebase.

Corana @ FM'19 [10] is a dynamic symbolic execution engine designed for multiple ARM Cortex variants. Complicated obfuscations such as packers, indirect jumps, and dead conditional branches (often referred to as opaque predicates) pose challenges to traditional approaches for malware analysis, including both static and dynamic methods. Corana is capable of effectively tracing IoT malware in the presence of these obfuscations. It was partly generated from natural language specifications of ARM instructions, and our paper demonstrates that this method can be systematically generalised to other variants, opening a potential research direction in applying formal methods to malware analysis.

References

[1] A. V. Vu, J. Hughes, I. Pete, B. Collier, Y. T. Chua, I. Shumailov, and A. Hutchings, "Turning Up the Dial: the Evolution of a Cybercrime Market Through Set-up, Stable, and Covid-19 Eras," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2020.

[2] A. V. Vu, D. R. Thomas, B. Collier, A. Hutchings, R. Clayton, and R. Anderson, "Getting Bored of Cyberwar: Exploring the Role of Low-level Cybercrime Actors in the Russia-Ukraine Conflict," in *Proceedings of the ACM Web Conference (WWW)*, 2024.

[3] A. V. Vu, A. Hutchings, and R. Anderson, "Yet Another Diminishing Spark: Low-level Cyberattacks in the Israel-Gaza Conflict," *Under Submission*, 2024.

[4] A. V. Vu, A. Hutchings, and R. Anderson, "No Easy Way Out: the Effectiveness of Deplatforming an Extremist Forum to Suppress Online Hate and Harassment," in *Proceedings of the IEEE Symposium on Security and Privacy (S&P)*, 2024.

[5] A. V. Vu, B. Collier, D. R. Thomas, J. Kristoff, R. Clayton, and A. Hutchings, "Assessing the Aftermath: the Effects of a Global Takedown against DDoS-for-hire Services," *Under Submission*, 2024.

[6] A. V. Vu, L. Wilson, Y. T. Chua, I. Shumailov, and R. Anderson, "ExtremeBB: A Database for Large-Scale Research into Online Hate, Harassment, the Manosphere and Extremism," in *ACL Workshop on Online Abuse and Harms (WOAH@ACL)*, 2023.

[7] I. Pete, J. Hughes, A. Caines, A. V. Vu, H. Gupta, A. Hutchings, R. Anderson, and P. Buttery, "PostCog: A Tool for Interdisciplinary Research into Underground Forums at Scale," in *Proceedings of the IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2022.

[8] L. Wilson, A. V. Vu, I. Pete, and Y. T. Chua, "Identifying and Collecting Public Domain Data for Tracking Cybercrime and Online Extremism," in *Open Source Verification in the Age of Google*, 2024.

[9] M. Tran, I. Choi, G. J. Moon, A. V. Vu, and M. S. Kang, "A Stealthier Partitioning Attack against Bitcoin Peer-to-Peer Network," in *Proceedings of the IEEE Symposium on Security and Privacy (S&P)*, 2020.

[10] A. V. Vu and M. Ogawa, "Formal Semantics Extraction from Natural Language Specifications for ARM," in *Proceedings of the International Symposium on Formal Methods (FM)*, 2019.