Google T&S Research Proposal: Marshall van Alstyne, Swapneel Mehta, Mayank Varia

1. Proposal Title: Illuminating the Shadows: Early Warning Systems for Emergent Online Financial Fraud and Scams
2. PI: Marshall van Alstyne, Questrom Professor in Management at Boston University (BU)
3. Co-PI / Project Lead: Dr. Swapneel Mehta, Postdoctoral Associate, BU; Research Affiliate, MIT
    a. Project Advisor: Mayank Varia, Associate Professor in the Faculty of Computing & Data Sciences at BU

4. Research objectives/Goals/Hypothesis

Data breaches and the consequent[1] fraud and scams have resulted in billions of dollars of losses for individuals with over 10 billion dollars[2] in losses from financial fraud and over 8 billion in scams[3]. With fraud the prime contributors are identity theft, carding, and other data breaches of so-called 'combolists' of user data. With financial scams, the attacks are directly targeting the users whose data have likely been breached. In comparison to the degree of enterprise social intelligence gathered from platforms, such topics are vastly understudied in academia. In order to benefit vulnerable populations and guide effective policymaking in this space, we propose to conduct scientific analyses of cybercrime networks on underground platforms and instant messaging applications using a combination of cutting-edge graph machine learning and open-source intelligence (OSINT) technologies in order to identify common patterns, preempt the harms arising from data breaches by developing early warning systems for online fraud and scams.

Research Question: What indicators can be extracted from cross-platform cybercrime intelligence data to identify the sharing patterns for breached data and develop early warning systems for financial fraud?

Goals:

1. Develop techniques to measure the cross-platform activities of cybercrime actors leaking user data and so-called 'combolists' of private user data, which result in downstream financial fraud and scams.
2. Identify coordinated networks of actors and campaigns engaging in fraudulent activity based on analysis of data breaches, data collection across platforms, and messages shared on publicly available groups.[4]
3. Publish a report analyzing financially fraudulent activity and the nature of networks promoting such activities across 3 platforms with a discussion of common patterns in order to support effective policymaking and mitigation.
4. Stretch goal: Contribute open-source tools to complement the detection of fraudulent campaigns on underground and encrypted platforms as well as relevant consumer-facing surfaces like instant and text messaging applications including a field test to validate its real-world utility by analysts and researchers.[5]

Deliverables: An analysis of cybercrime actors across 2 underground marketplaces and public channels across 15 public groups within 1 instant messaging application with robust data collection mechanisms that operate in Android ecosystems, within platform terms of service, advancing what past work has demonstrated.[6,7,8]

We identify two types of platforms–messaging apps and online forums–and conduct our analysis across multiple platforms so that our work will provide a robust template for extension into examining cybercrime networks in similar

[1] Phishing and Extortion Scams After Security Breaches
[2] Facts + Statistics: Identity theft and cybercrime | III
[3] New FTC Data Show Consumers Reported Losing Nearly $8.8 Billion to Scams in 2022
[4] TGStat - List of Public Telegram Channels by Country and Topic
[5] Understanding Telegram's ecosystem of far-right channels in the US | by @DFRLab
[6] GitHub - ElDavoo/wa-crypt-tools: Decrypts WhatsApp .crypt12, .crypt14 and .crypt15 files.
[7] Tools for WhatsApp Data Collection
[8] Code to get data from WhatsApp public groups

ecosystems. We aim to support the academic community by releasing software for academic research that is useful for tracing cross-platform cybercrime networks. We optionally include a field test via industry partnerships including the potential testing of our system by external analysts to support their investigations as we have done previously working with The Sunday Times[9] and Deutsche Welle Akademie, among others.

5. Proposed approach

Build effective multilingual alert capabilities for emerging data breaches and financial cybercrime threats. Set of proposed activities involved in the proposed approach:

1. Cross-platform information gathering - Develop APIs and scrapers to correlate actors, postings, product listings, ratings, and social graphs across both dark web and surface platforms to identify intersections signaling coordinated fraud.
2. Multilingual understanding - Utilize language translation services and develop classifiers trained on annotated non-English financial cybercrime text across forums/markets to enable monitoring beyond English.
3. Narrative detection - Employ natural language techniques like topic modeling, named entity recognition, and semantic analysis to systematically identify shared narratives, terminology, and concepts indicative of emerging threats.
4. Anonymized data sharing - Use privacy-preserving means to access and analyze platform data from end users to identify prevalent fraud targeting end users following past work into Whatsapp and Telegram monitoring.
5. Stretch Goal: Alert testing and iteration - Work closely with financial industry partners to test and refine the performance of generated alerts against known fraud to track known patterns, improve precision, and lower false positives.
6. Stretch Goal: Responsible disclosure - Coordinate with law enforcement and CERTs to ensure emerging threats are handled appropriately once detected through a responsible disclosure policy.

6. Expected Impact/Benefit to research community

Online financial fraud and cybercrime pose growing threats to consumers in the digital age. Scams, phishing campaigns, and the sale of leaked personal data on underground forums often lead to substantial monetary losses and privacy violations. The ubiquity of encrypted messaging platforms has also enabled new techniques for information laundering[10], allowing harmful actors to more covertly spread leaked credentials before they become public. Misleading information on social networks furthers the reach of financial scams.

These dynamics disproportionately impact marginalized groups, especially those with lower financial literacy and digital access[11]. More concerningly, existing monitoring systems struggle to keep pace as schemes grow more sophisticated. There is an urgent need for enhanced techniques to identify emerging cybercrime networks and preempt widespread financial harms through early detection and alerts. More research grounded in robust technical frameworks, data-driven insights, and transparency is essential to guide consumer protection policies and regulation of encrypted ecosystems in the EU's Digital Services Act.

7. Budget: Amount requested, inclusive of all fees (*FTE = full-time equivalent)

Name	Units	Description	Total (USD)	FTE

[9] Building Parrot Report and running disinformation tracking workshops for DW Akademie Staff

[10] Encrypted messaging apps are the future of propaganda | Brookings

[11] Serving Communities of Color | FTC Report

Research Assistants	5	Support Software Development and Platform Data Collection Infrastructure at scale for 12 months.	57500	0.5
Cloud Infrastructure	1	BigQuery, Cloud Storage, Compute Engine VM Instances, Vertex AI Notebooks. Cost based on a similar stack for Parrot Report for 12 months on Google Cloud Platform.	16000	
Project Lead	1	Project management and coordination of research activities. Presenting and analysis report writing.	16500	0.2
Equipment	5	Laptops/Mobile devices for testing apps and data collection	7500	
Travel and Publication	2	Conference travel; presentations for feedback from academic and industry stakeholders	2500	

Project Budget: USD 100,000

8. Disclosure Policy:
To maximize impact and enable further research, we intend to openly share key outputs from this project including publications, code, and anonymized threat data. Our publications detailing the analysis framework, models, and findings will be submitted to leading security conferences and machine learning for security venues. Our team is actively engaged in this field. If time permits, we aim to publish our methodology and design software interoperable with other cybercrime monitoring tools and databases so that it enables the academic community to review, extend, and reuse the tools that we develop. Details of our models and classifiers will also be published to facilitate reproducibility. Finally, if within future platform terms of service, we will host anonymized multilingual datasets containing samples of narratives, posts, and other signals labeled for financial cyber threats on platforms like Kaggle, FigShare, and Data.World to encourage broader study into identification of emerging multilingual signals and harms.

9. Expected results/publication with timeline

Weeks 1 - 10: **Identify Platforms and Track Digital Trace Data**
Track and collect data from newly emerging forums, encrypted instant messaging applications, marketplaces, and online communities focused on trafficking stolen financial information and user credentials. This involves monitoring surface, deep, and dark web channels for new platforms, creating mechanisms for user-data sharing and platform data donations research from target user populations in the regions of interest. With necessary approvals, we aim to collaborate with industry partners to identify relevant channels and use known data breach tracking platforms, striving to ensure interoperability with related tools, data breach, and vulnerability databases to advance an interoperable toolkit.

Weeks 12-16: **Setup Multilingual Translation Infrastructure for non-English Data**
Analyze non-English (Spanish, French, German) conversations around financial cybercrime to uncover region-specific threats that are untracked by publicly available analytics tools with a minimum of one additional language to English. Advance existing techniques to correlate actors, postings, and conversations across both newly identified and previously known platforms to uncover intersections between different cybercrime ecosystems. We will require in-house LLM-based language translation such as using M2M100 12B due to the prohibitive costs of using external services including Google Translate, DeepL Pro, deep-translator which we have already explored at scale for YouTube comment translation.

Weeks 18 - 36: **Narrative Detection**

Identify narratives by using multilingual classifiers to flag conversations about illegal or terms of service-violating activity such as the attempted sale of new credential dumps, database breaches, or sharing of exploitation tools that could enable future data theft. Track longitudinal shifts in narratives, pricing models, platform usage, and other indicators that may signal evolution of capabilities and cybercrime business models.

Weeks 40 - 52: **Report Writing and Deployment**
Complete writing and submit the publication at a relevant cybersecurity-focused academic venue and optionally submit talk proposals at industry-focused venues where we have previously presented, such as mWise and DEFCON, for feedback from industry experts. Deploy the system to auto-generate early warning alerts when a new breach, credential dump, or exploitation tool release is emerging across platforms. If time and resources permit, we can test the system's efficacy for a 4-week period over multiple threat activities across platforms. We hope to contribute multilingual datasets for training systems to identify new financial cybercrime threats emerging in underresourced non-English languages.

Anticipated Challenges: Data sharing by social media platforms is undergoing a paradigm shift as platforms move to protect their data and prohibit research data access. Being mindful of the costs of platform data shutdowns and their downstream delays on research projects like ours, we rely primarily on external, privacy-preserving data collection mechanisms that don't violate platform terms of service. For this reason, while we are likely able to deliver on the stretch goals such as a platform to collect cybercrime datasets and networks, we remain conservative in listing grant deliverables and include 2-4 weeks as buffers in the above timeline.

10. Statement of prior work, including a list of relevant publications.
Given the page limits we provide a GDrive containing the detailed curriculum vitae of the applying team with relevant highlights summarized below:

Marshall van Alstyne, Allen and Kelli Questrom Professor in Information Systems and Professor, Information Systems, Questrom School of Business at BU.

I am one of the world's foremost experts on network business models and coauthor of the international bestseller 'Platform Revolution'. I conduct research on information economics, covering such topics as the economics of speech markets, platform economics, intellectual property, social effects of technology, and productivity effects of information. I have been a major contributor to the theory of two-sided networks taught worldwide, and to the theory of platforms as inverted firms, applied in antitrust law. My work includes a top 50 all-time article for Harvard Business Review, and won the INFORMS IS 2020 and Herbert Simon 2021 awards for research with real world impact.

Contributions to this Project: Policy and Governance implications; Model incentive structures and market dynamics for cybercrime actors, identify platform governance changes to mitigate harms, develop policy advisory insights for EU regulation, and a formal framework to understand cybercrime activities across platforms.

Relevant work:

1. The EU Digital Markets Act A Report from a Panel of Economic Experts: highlights a template to draw insights from the upcoming EU Digital Services Act and UK Online Safety Act on the topics of cybercrime threat actors which our research will provide evidence for platforms to track and protect users against.

Past work into markets and misinformation, modeling the complex system of actors participating in online information sharing, collectively will help identify patterns, systemic vulnerabilities, and policy loopholes exploited by threat actors like cybercriminals: see A Market for Truth to Address False Ads on Social Media; Free Speech and the Fake News Problem; and Platforms and Ecosystems: Enabling the Digital Economy.

Swapneel Mehta, Postdoctoral Associate at BU and Research Affiliate at MIT

Google T&S Research Proposal: Marshall van Alstyne, Swapneel Mehta, Mayank Varia

I'm a recent Google Research Innovator and received my Ph.D. at NYU Data Science and their Center for Social Media and Politics researching methods to limit online disinformation on social networks using techniques from machine learning and causal inference. My dissertation research focuses on improving measurement methodology to identify the effects of interventions, and in deploying auditing tools for online harms in order to improve platform governance. I conducted research with Oxford and Meta, previously interned with Slack, Twitter, Adobe, and worked on machine learning for particle physics at CERN. My dissertation includes the referenced work on analyzing social media interventions made available (internal-only) at this link for grant review purposes.

Contributions to this Project: Data collection and analysis from underground platforms and instant messaging applications, machine learning models for multilingual translation and multimodal analysis, project coordination and leadership, supervision of research assistants.

Relevant work:

1. Estimating the Impact of Coordinated Inauthentic Behavior on Content Recommendations in Social Networks: Modeling the effects of adversarial activity attempting to game recommender systems with millions of real-world network data points collected from Reddit.
2. Open-Domain Trending Hashtag Recommendation for Videos: Building robust recommender systems at scale.

I founded a nonprofit research collective called SimPPL to create open-access trust and safety tools working with student communities in the global south. We are actively deploying trust and safety tools like https://parrot.report (for the Sunday Times) that identifies coordinated networks promoting unreliable Russian media online on Twitter using quantitative metrics interpretably scored by graph algorithms and machine learning. We're also building open-source, cross-platform tools to conduct network analysis on YouTube, Whatsapp, Telegram, and fringe-social platforms with existing expertise and impact discussed in these slides.
Our past work has a strong and clear overlap with the proposal, positioning us well for developing robust data collection mechanisms for analyzing understudied cybercrime networks to accelerate academic research into this space. SimPPL has worked with The Sunday Times, Deutsche Welle, and others, and have won grants and awards from Google Research India, Google Cloud, Amazon AWS, the Wikimedia Foundation, the Goethe Institute, and the NYC Media Lab, and is well-placed to work with partners for a potential field test of the open-access technology proposed to be developed.

Mayank Varia, Associate Professor in the Faculty of Computing & Data Sciences at BU.

My research explores the computational and social aspects of cryptography, including the design, development, and deployment of privacy-respecting systems using cryptographically secure multiparty computation to protect data while in use. My designs for accessible, equitable, and socially-responsible data analysis have been used to determine the gender wage gap, workplace culture in the museum industry, and repeat offenders of sexual assault in universities (inspired by the #MeToo movement). Additionally, I serve on the United Nations Privacy-Preserving Techniques Task Team and served on the United States Advisory Committee on Data for Evidence Building to promote the use of cryptographically protected data analysis and shape the laws and policies surrounding its use.

Contributions to this Project: Secure Data Analysis for End-to-end encrypted platforms, Privacy-preserving algorithms for collecting and analyzing private data, Can WhatsApp Messages Be Secure and Encrypted—but Traceable at the Same Time? | The Brink | Boston University

Relevant work:

1. Hecate: Abuse Reporting in Secure Messengers with Sealed Sender: a cryptographic method for abuse reporting in end-to-end secure messengers, published at USENIX Security 2022. I am actively working on extensions to this mechanism that would support mis- and disinformation reporting.

Other relevant work includes Universally Composable End-to-End Secure Messaging: A Modular Analysis; and SoK: Cryptographically Protected Database Search.