EECS 349 & 444: Group Project Announcement

- **Group project:** In the project, you are required to use cutting-edge techniques to solve the proposed cybersecurity research problems.
  - 3-4 students per group
  - Select a seed idea for your group project
  - Fully motivate the problem and survey related work (10%)
  - Project preparation (e.g., data collection, annotation, preprocessing, surveys, etc.) (20%)
  - Develop your own solutions - substantial novel technique development and implementation (20%)
  - A thorough empirical evaluation and validation (15%)
  - A fully developed project report (20%): **You should NOT copy anything from anywhere!!**
    EECS 349: 8-page in ACM Master article template for LaTeX
    EECS 444: 12-page in ACM Master article template for LaTeX
    https://www.acm.org/publications/proceedings-template
    (Note: if your team is a mix of EECS 349 & 444, then your team needs to submit 12-page report)
  - Project presentation (15%): 12-min presentation + 3-min Q/A

T1: Gaining Deep Insights into the Online Underground Ecosystem

Motivation and Background: Nowadays, many sophisticated underground markets (e.g., underground forums, social media groups) have emerged over the Internet, where cybercriminals exchange information with fellow criminals on abusive tactics and engage in the sale of illicit goods and services. The function for these markets is not only for social contact within users, but also to support criminal activities, such as buying or selling crimeware such as malware and crimeware-as-a-service (CaaS) such as hacking services. In order to allow law enforcement communities to devise effective disruptive strategies, there's an urgent need for novel techniques and tools to gain valuable insights into these underground markets.

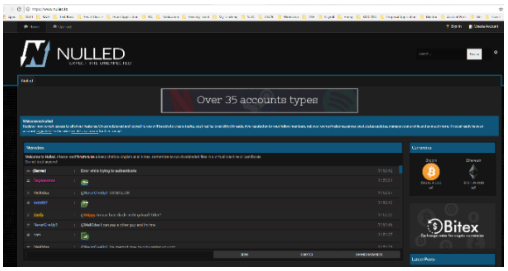

Figure 1. Nulled Forum

To gain deep insights into underground markets and better understand the cybercrime ecosystem, in this project, you are asked to first find at least five active underground markets, such as underground forums (e.g., ***Nulled***: https://www.nulled.to/#!Marketplace, ***Hack Forums:*** https://hackforums.net/forumdisplay.php?fid=107). You need to describe and summarize how you find these underground markets and explain why these markets are significant for gaining insights into the online underground ecosystem. Then conduct the following research tasks:

1. Select one underground market you explore, based on which each group needs to focus on at least one particular kind of crimeware (e.g., exploits, botnets) and one kind of CaaS (e.g., malware attack, hacking service) traded in the markets. For each type of crimeware (denoted as P1) or CaaS (denoted as S1), you need to first develop your own solutions/tools to collect a number of threads (>50 threads for P1 and >50 threads for S1) and their related comments (>30comments/thread) for further analysis (see steps 2-3). Describe and summarize how you collect the data. *Note: we only collect and analyze those threads whose comments > 30.*
2. Base on the collected data, you are asked to develop your own solutions/tools to analyze: (1) crimeware/CaaS trading **threads**: i.e., extract username and profile of vendor, product/service name of each thread, price, payment method, # of comments (i.e., replies), # of reviews; (2) **comments**: i.e., classify each comment (i.e., username and profile of commenter, trading [Yes/No/Uncertain], contracted customer [Yes/No], review [Positive/Negative/Neutral], Q&A [Yes/No], other). You need to submit the analysis results using the required template shared in Canvas. Your analysis/annotation will be validated by cross-validation during grading.
3. Based on the above steps, please develop your own solutions/tools for in-depth analysis: (1) find out the **key players** (i.e., most active vendors and buyers) for the kind of crimeware (P1) and CaaS (S1) you explore; (2) further analyze the top key players (one vendor and one buyer) for P1 and S1 to find out: i) whether he/she is an individual or organization; ii) what other products he/she sell or buy; iii) how he/she influent others in the market; iv) if he/she is active in other markets and how he/she will have the impacts in the cyberspace, etc. Describe the storyline and provide the case studies to elaborate your findings.
4. Based on the above findings and analysis, devise your solutions to help inform effective countermeasures.
5. A fully developed project report with required format should be submitted.
6. Finally, present your project in the class.

T2: Study code security problem in social coding platforms

Motivation and Background: Unlike conventional approaches, modern software programming cyberinfrastructure, consisting of online programming discussion platforms (e.g., Stack Overflow) and social coding repositories (e.g., GitHub), has offered an open-source and collaborative environment for distributed scientific and engineering communities to expedite the process of software development. Within the ecosystem, researchers and developers can reuse code snippets/libraries or adapt existing ready-to-use software to solve their own problems (e.g., CycleGAN in GitHub). Despite the apparent benefits of this new social coding paradigm, its potential security-related risks have been largely overlooked: as shown in Figure 2 and Figure 3, insecure or malicious codes could be easily embedded and distributed (e.g., through copy-and-paste), and these codes can also be deliberately disseminated for production software. The goal of this project is to detect the malicious projects hosted in GitHub and analyze how these threats are propagated/disseminated in the cyberspace.

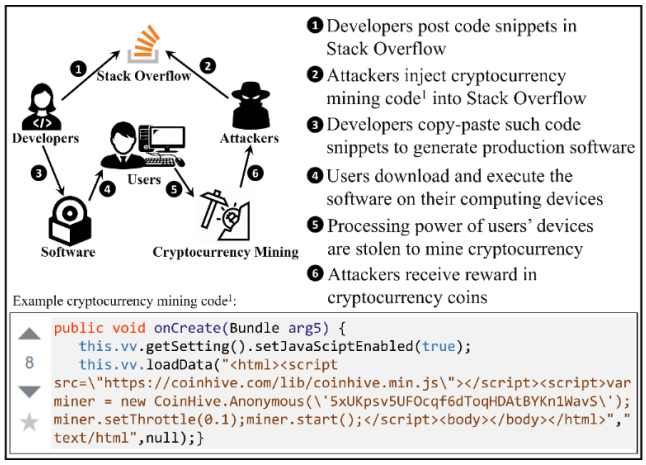

Figure 2. Example of code security attacks in Stack Overflow.

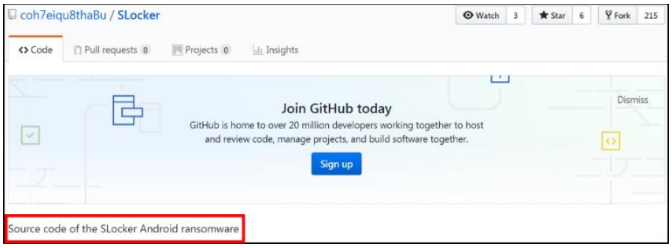

Figure 3. Android ransomware hosted in GitHub.

1. You are first asked to explore and devise your own methods to identify the malicious codes/projects hosted in GitHub. Describe and summarize how you identify them. Note: you can regard the code/project is malicious if over 1/4 of AV engines in VT detected it as malicious.
2. In order to leverage the pre-identified samples to automate the detection of malicious codes/projects in GitHub, you should propose your own solutions and assess the effectiveness of your developed tools. You need to submit the ground truth using the required template in Canvas (each group needs to detect > 100 malicious projects hosted in GitHub). Your submission will be validated by cross-validation during grading.
3. Based on the detected malicious codes/projects hosted in GitHub, you are asked to develop your own approaches and tools to further analyze how these threats are propagated or disseminated in the cyberspace: (1) Who are the key players? What is his/her profile, his/her reputation, geo-location, etc? (2) How they disseminate the malicious codes/projects hosted in GitHub and interact with other users cross-platforms (e.g., between Stack Overflow and GitHub to propagate or disseminate the malicious codes/projects)? Afterwards, please further analyze and figure out their dissemination networks (i.e., the communities hiding behind if any) and how they interact with each other. You need to find out at least 10 solid cases to demonstrate how the malicious codes/projects are disseminated in the cyberspace. Describe the storyline and provide the case studies to elaborate your findings.
4. Based on the above findings and analysis, devise your solutions to help inform effective countermeasures.
5. A fully developed project report with required format should be submitted.
6. Finally, present your project in the class.