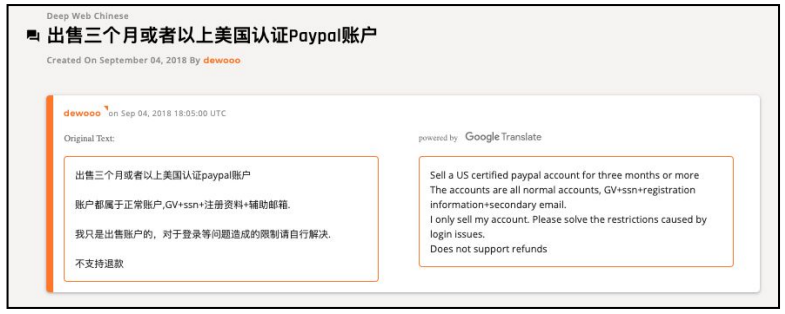
"Chinese Cyber Crime: A Graph Approach"

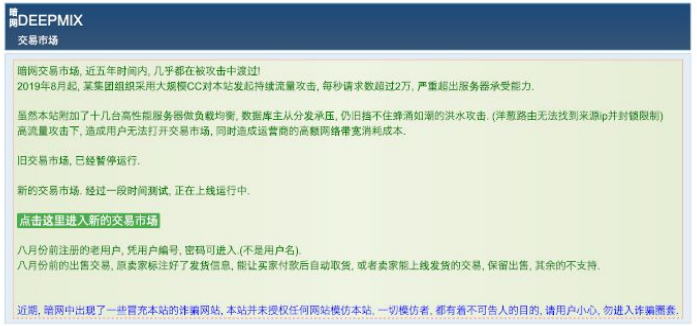
Aaron Shraberg

Why do this?

- **User-friendly**
  - Many tools out there aren't intuitive or designed to specific use cases
- **Intuitive**
  - Visualizations can aid in analysis, threat landscaping
  - Data models can be calibrated as our understanding changes
- **Threat Tracking**
  - Data is historical but is also useful for understanding amorphous cyberunderground, continuity of threat actors and threat landscaping.



DEEPMIX (暗网交易市场) - 9/4/2018

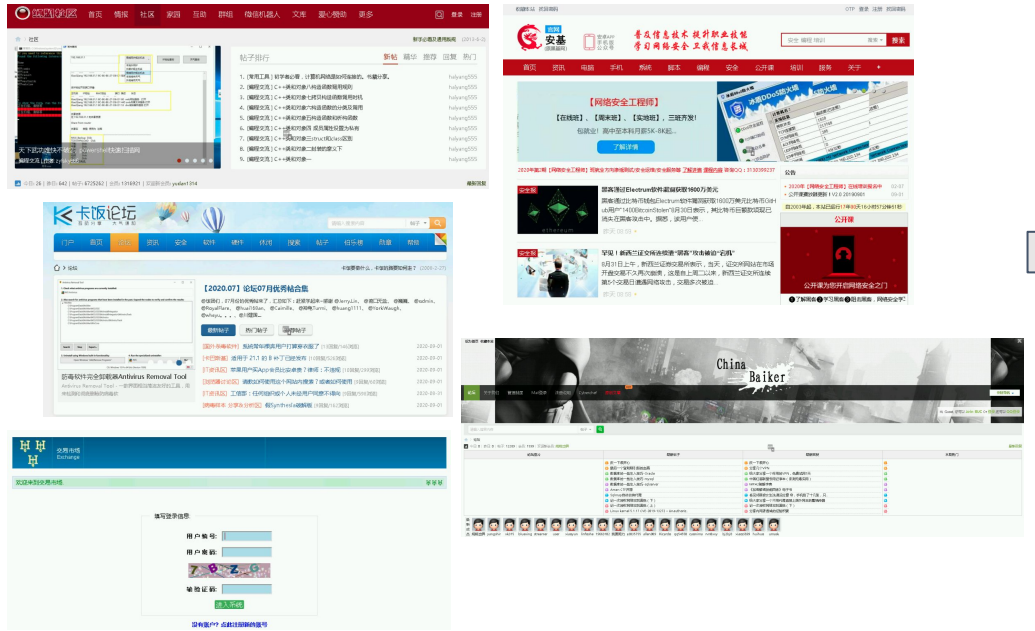


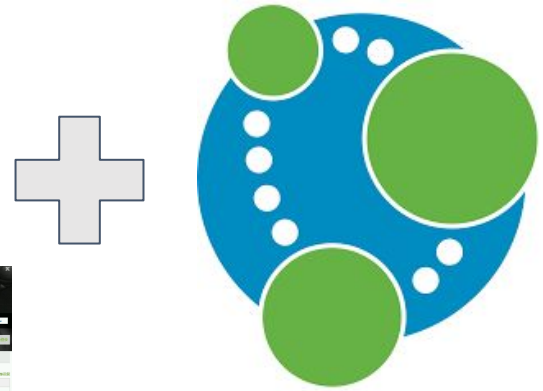
Notice of Closure and Reopening ~ 9/2019

Exchange (交易市场) - 2/21/2020

What was done?

- Thirty-five Chinese language websites were populated into a Neo4J graph instance

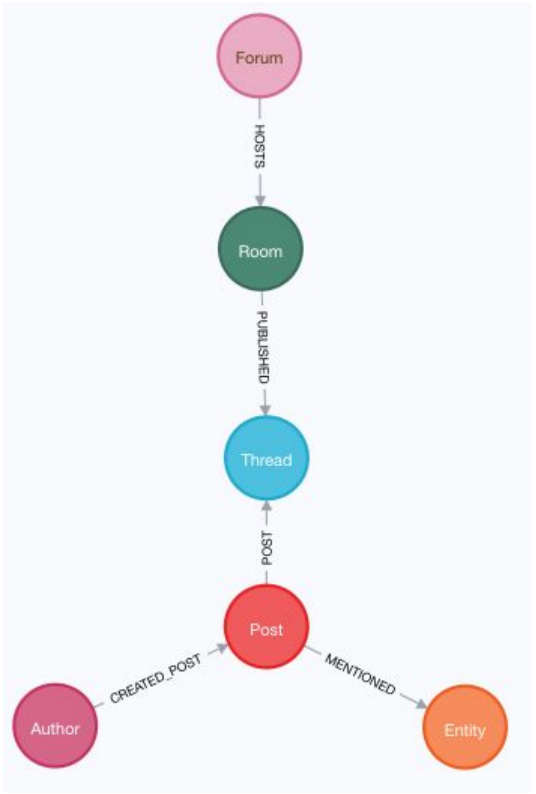




?

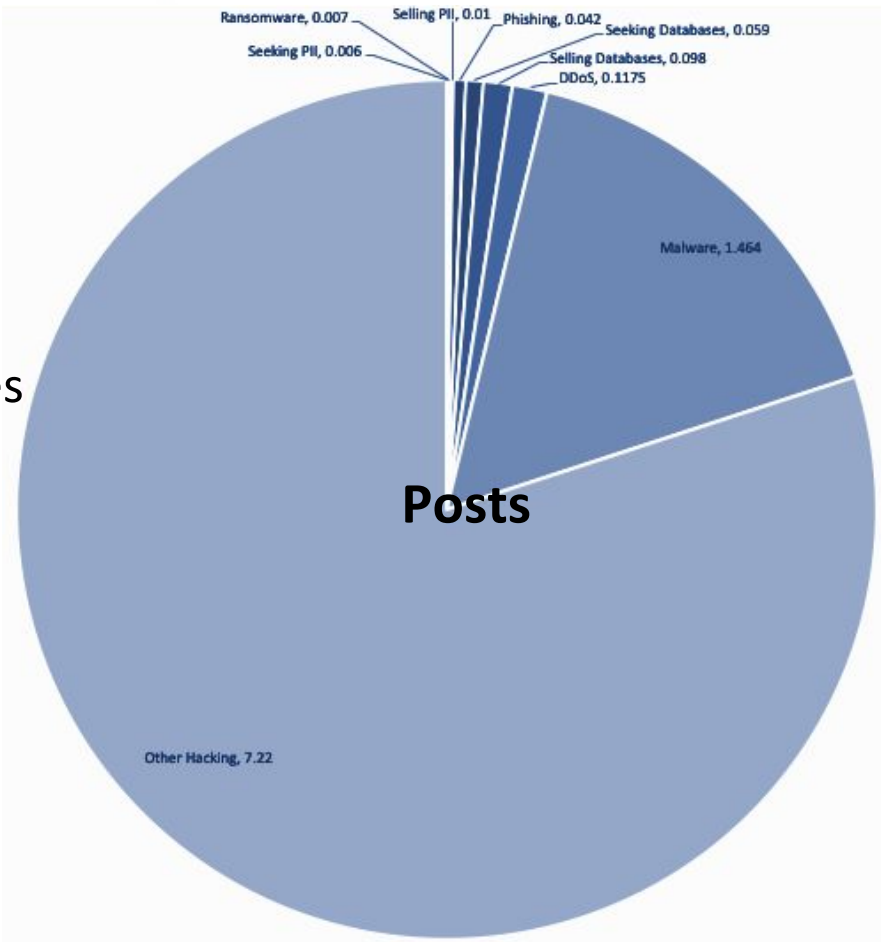
By the numbers

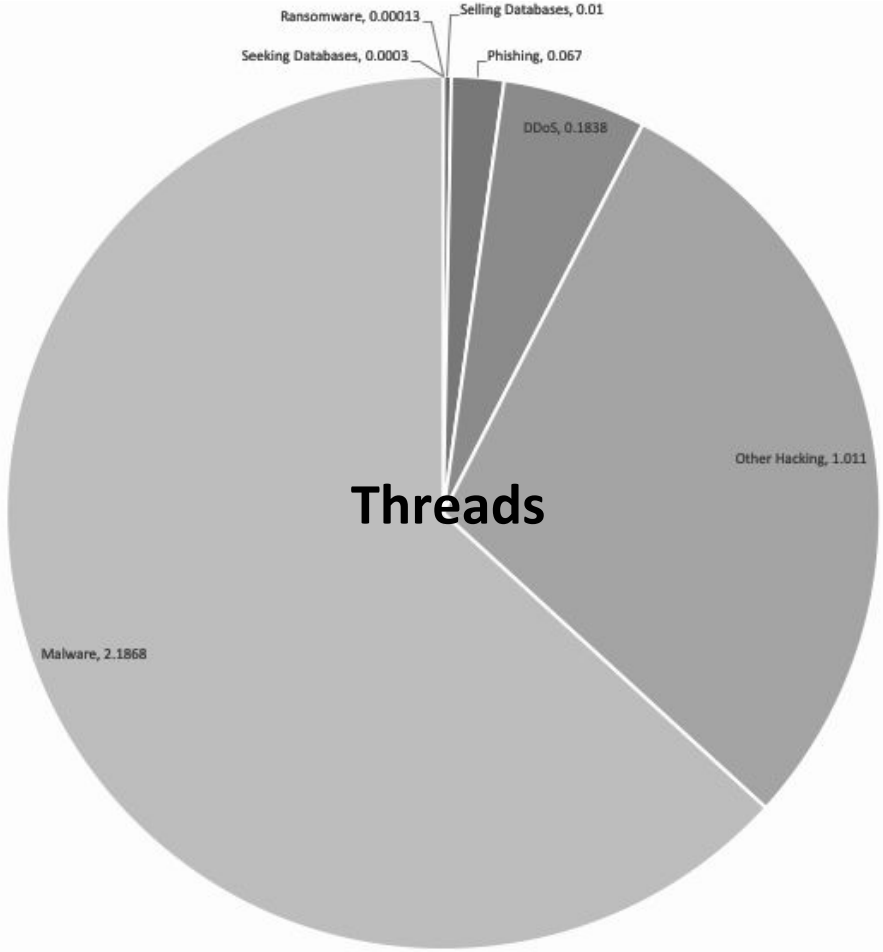
Rooms	**1,614**
Threads	**1,272,439**
Posts	**22,611,431**
Aliases	**1,940,565**
Entities	**1,495,495**



By the numbers

- 1st Malware
- 2nd DDoS
- 3rd Selling Databases
- 4th Seeking Databases
- 5th Phishing
- 6th Selling PII
- 7th Seeking PII
- 8th Ransomware



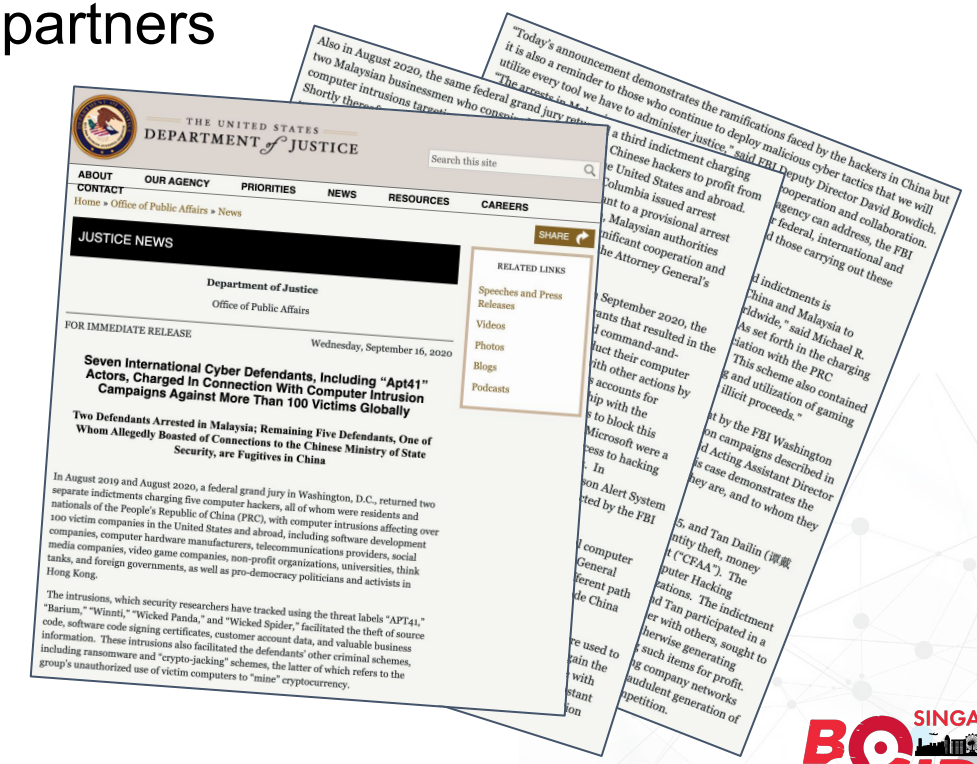


Slang

Slang	Meaning	Explanation
"Pants"；裤子	Database	The Chinese word for pants sounds like the Chinese word for database, also pronounced "kù"
MM	Trojan	M is the first letter of both words for trojan when spelled using the romanization (**mùmǎ**；木马)
Yuan控	RAT	Romanization of Chinese word for "remote" ("yuǎn"；远) combined with Chinese character for "control" (控)
D单	DDoS	D is the first letter of **D**DoS followed by the Chinese character for "single," as in "single point of congestion."

Findings - Ransomware

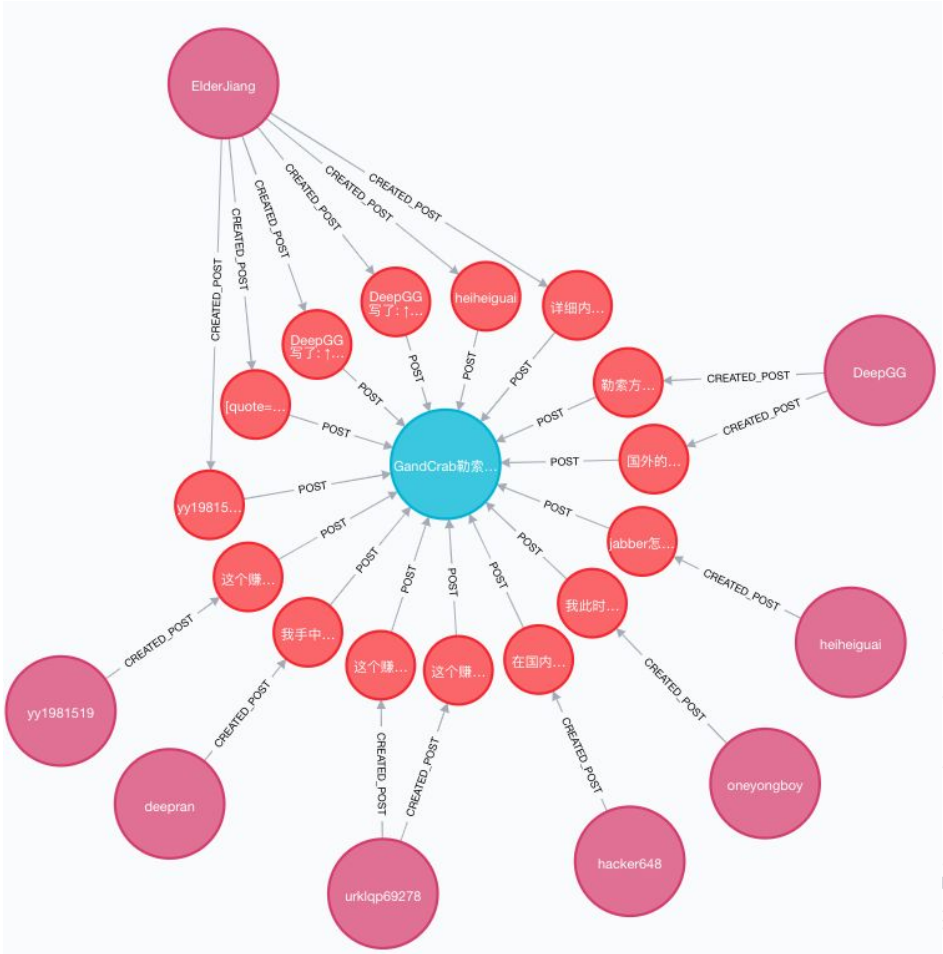
- Just 1,595 posts mention ransomware, accounting for 0.007 percent
- 300 threads had ransomware in topic name, accounting for 0.00013 percent
    - Small but impactful dataset
    - Recruitment for ransomware partners



Findings - Ransomware

- Chinese cyber crime actors have used forums to recruit for ransomware gangs

- **Conclusions:**
  - Recruitment ongoing however it appears to have shifted to **chats**

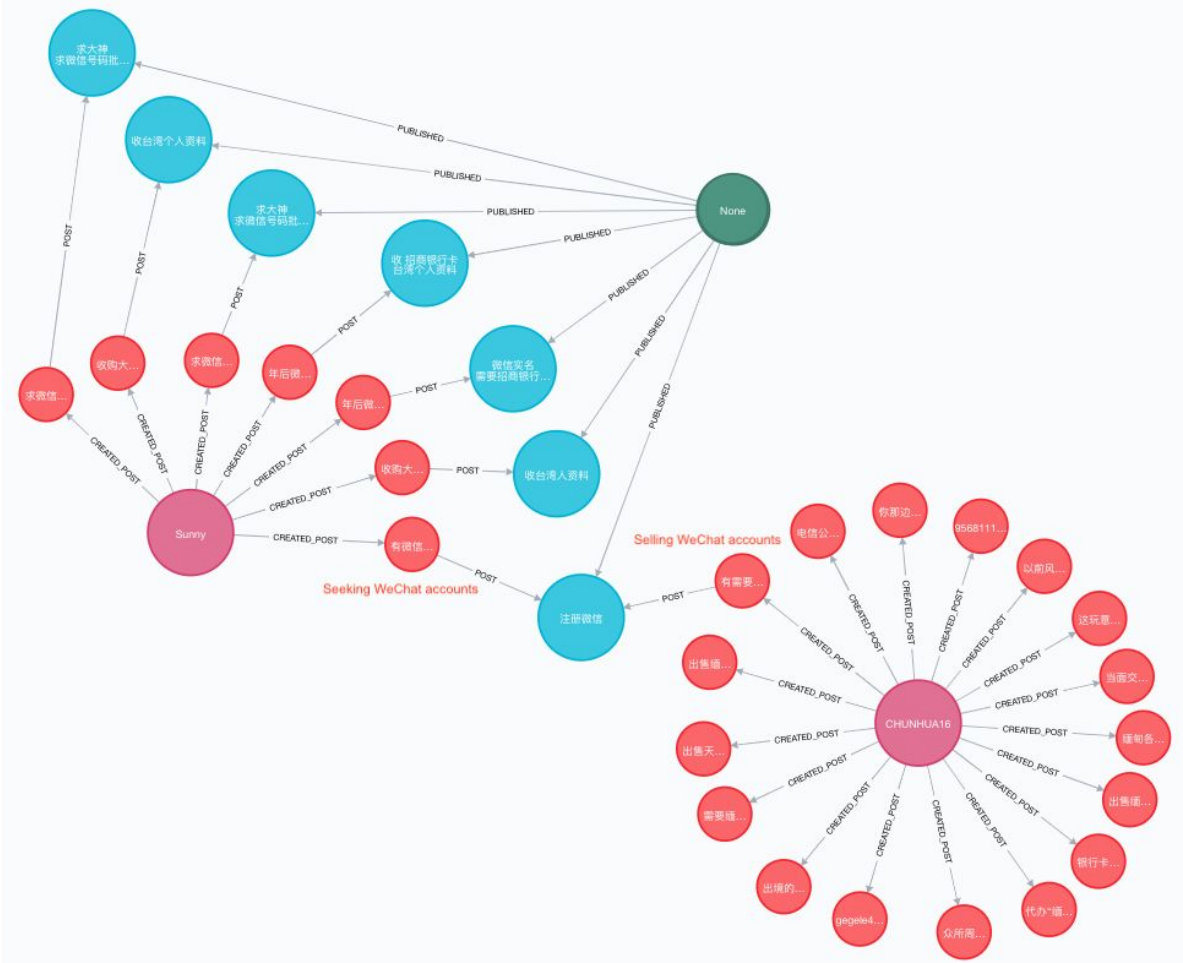


Findings - Seeking PII

- A relatively small number of threat actors seeking personally identifiable information (PII)
    - 0.006 percent or about 136,000 posts out of 22.6 million
- Pivoting off of one thread or post about seeking PII, however, revealed a broader network of posts about other fraud activities that may leverage PII.

Findings - Seeking PII

- "Sunny" seeks WeChat accounts for Taiwan
- "Chunhua" has WeChat registrations for Myanmar
    - Significant chatter around WeChat registration techniques
- **Conclusions:**
    - Seeking PII activity relatively small scale, but it can have an outsized impact -- Taiwan
    - Investigations begin narrowly defined can (and should?) lead to further insights -- example techniques used to procure tools used by TAs

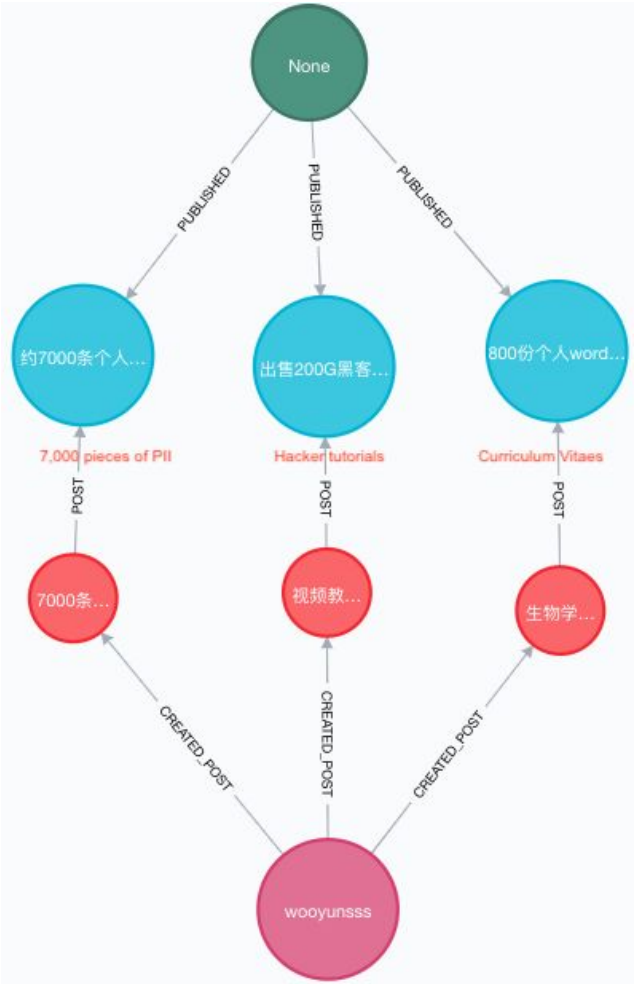


Findings - Selling PII

- A relatively small number of threat actors selling personally identifiable information (PII) within the sample dataset
    - 0.01 percent or about 226,000 posts out of 22.6 million
    - Twice the number of posts regarding seeking PII
        - Seekers are targeting specific types of data while sellers want to sell
- Pivoting off of one thread or post about selling PII shows multitude of advertisements by a seller

Findings - Selling PII

- "wooyunsss" selling "about 7,000 'pieces' of PII"
  - Other Sale items: 200 gigabytes worth of hacker tutorials, resumes focused around the biology field
- **Conclusions:**
  - Threat actors are pragmatic and sell a diversified list of products and services. Marketplaces become a veritable thrift shop of stolen goods and hacker services for sale.



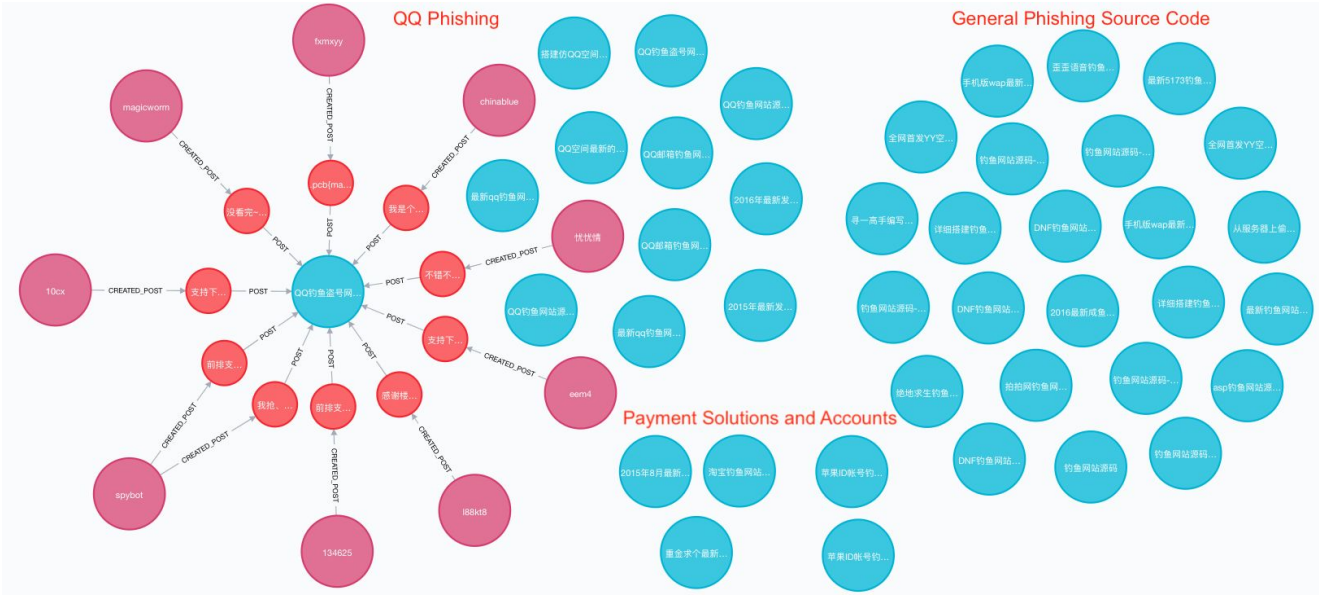
Findings - Phishing

- 9,520 posts and 854 threads mention phishing pages and phished emails, accounting for 0.042 percent of the total posts and 0.067 percent of the total threads.
    - Phishing is a often seen initial attack vector and it would come as little surprise to see an abundance of phishing-related TTPs being shared and discussed.

Findings - Phishing

- QQ, general source code, payment platforms and accounts

- **Conclusions:**
  - High interest in source code
  - Phishing source code is both targeting domestic QQ users and general source code that can be used domestically or abroad
  - Discussions are unambiguous



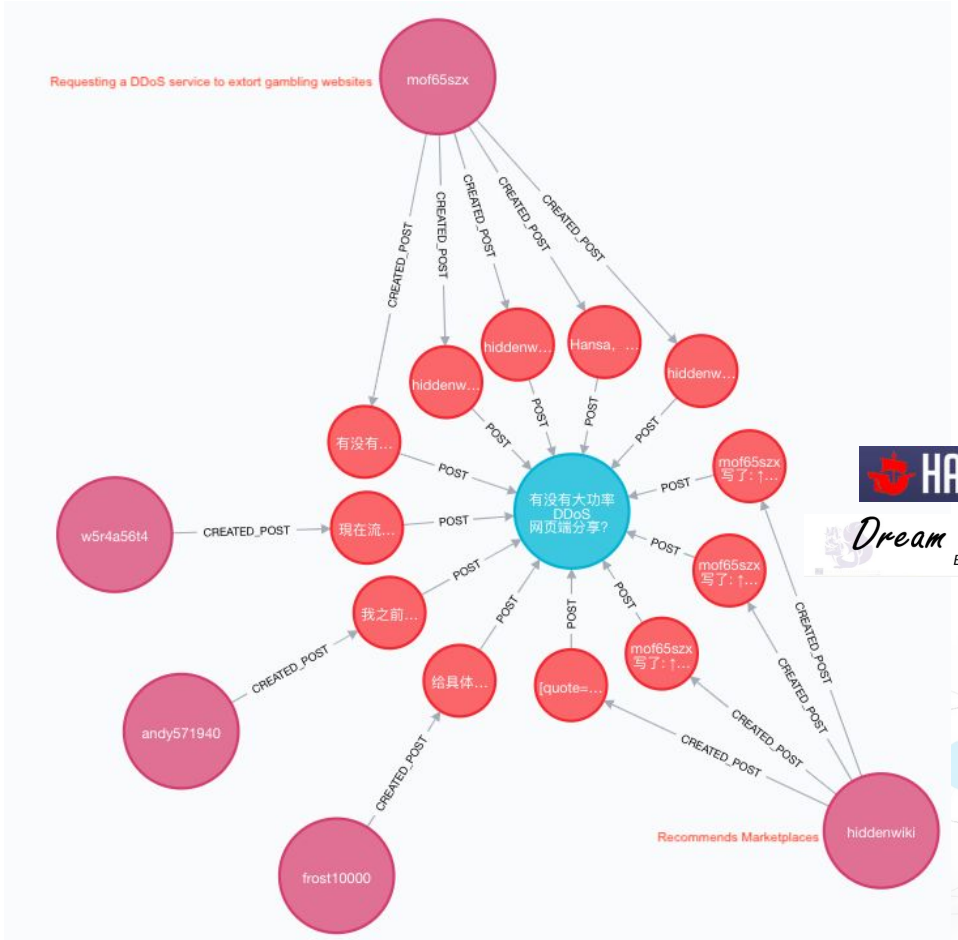
Findings - DDoS

- 26,581 posts mention DDoS tools, bots or zombies, accounting for 0.1175
- 2,339 threads mention DDoS tools, bots or zombies, accounting for 0.1838 percent of the total threads.
    - These posts and threads include DDoS dashboards, source code, and DDoS-for-hire services.

Findings - DDoS

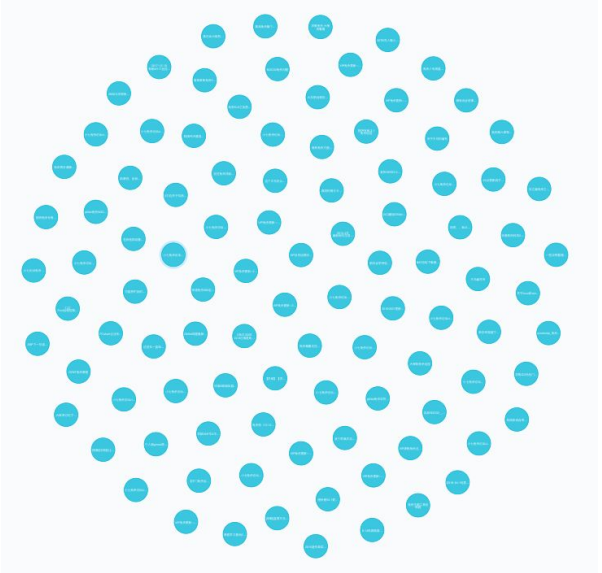
- There are many, many discussions around DDoS during this time (2006-2019).

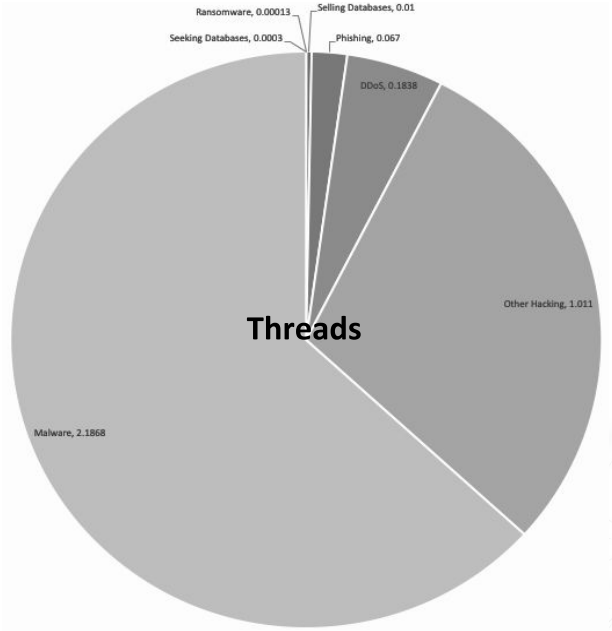
- **Conclusions:**
  - After malware, DDoS-related chatter is highest volume in the underground
  - Large amount of sharing in DDoS tools
  - DDoS is advertised as both stressor or weapon (e.g., "Great Cannon")



Findings - Malware

- 331, 250 posts mention malware, accounting for 1.464 %
- 27, 826 threads mention malware, accounting for 2.1868 %
    - Largest category
    - Includes trojans, anti-virus avoidance techniques, and malware





Findings - Malware

- VIP versions
- CVE Discussions
- Tools, Tactics and Techniques
- **Conclusions:**
    - Forums offer malware versions to privileged members
    - CVE Discussions can be tracked and used for triaging
    - Other TTPs discussed and shared

Research Limitations

- **Size unknown and constantly changing**
  - The data is a portion of an unknown number of hacking websites
  - Several Chinese DDW sites temporarily suspended registrations, gone offline, or rebranded under new URLs
- **Language is always changing**
  - Many terms signify the same threat type
  - Slang falls out of use while new terms appear

Key Takeways

- **What is the Chinese Deep and Dark Web and how big is it?**
  - Chinese speaking threat actors have had their forums and marketplaces gradually taken down
  - Hacker sites, reserved for hobbyists and cyber professionals, are disappearing
  - There's still some value there
- **Chats (and Tor!) are emerging as the primary view into threat actor behavior**
  - Fraud activity now emerging in chats
  - However, chat data is voluminous noisy and difficult to model
- **An approach that integrates desktop and mobile channels is key**
  - Evidence to suggest threat actors still use the old sites so they will continue to be relevant
  - Entity extraction: links include chat account numbers, emails, monikers and others…
- **Continuously discover new sources in traditional mediums**
- **Develop solutions to connect web-based and emerging channels of communication to optimize threat identification and horizon scanning**