Extracting cyber threat intelligence from social media with case studies in Twitter and Reddit

Dainora Jakstaite[1 [0000−1111−2222−3333]] and Ricardo M. Czekster[1 [0000−0002−6636−4398]]

Aston University, Birmingham, B4 7ET, United Kingdom
{149121837,meloczer}@aston.ac.uk

Abstract. A substantial amount of Internet traffic pertains to social media. This virtual way of interacting among peers, friends and audiences has contributed to help users test arguments and advertise statuses whilst helping organisations reaching out prospective clients. Given its global outreach, social media users produce a massive amount of daily data that are publicly available (depending on T&C). This brings interesting challenges in Cyber Threat Intelligence (CTI) as it can employ social media datasets to investigate impending cyber-attacks or latest malicious incursions on victims. Our contributions are two-fold, firstly, it outlines how to extract CTI in a timely fashion as well as the considerations and trade-offs for cyber security officers with case studies in Twitter and Reddit. Secondly, it proposes a *trustability metric* for determining reputable accounts to prioritise subsequent analysis efforts.

Keywords: Cyber Threat Intelligence · Cyber security · Metrics for timely decisions · Social Networks Analysis.

1 Introduction

Modern society has become increasingly dependent on social media outlets as a *de facto* communication and entertainment venue to reach out family, friends, and communities, be informed on latest events, and interact with others. Previous years have witnessed the establishment of social media in general as a conduit for letting people and organisations promote business, inform audiences, or congregate users around similar preferences and points of views. The now large community of adopters aggregated around these networks enabling researchers to tackle its inherent benefits with Social Networks Analysis (SNA) [21] as a valid alternative to analyse complex systems in social spaces.

Latest years have seen the development of a myriad of social media instances such as Facebook, Twitter (re-branded to 𝕏 in July/2023), YouTube, Reddit, LinkedIn, Instagram, TikTok, and more recently, Threads, to mention a few. Given its requirements to sustain a global outreach, there are substantial technological challenges for keeping up, update, and maintain security whilst providing a privacy preserving user experience. The sheer scale of data produced daily in social media makes it a good target for conducting timely analysis that

could help managers better understand interactions and information dissemination over large networks. That is the main reason to consider social media data streams as a crucial element in analysis [20,8].

From a cyber security perspective, they may provide invaluable data to help deter cyber-attacks before they spread over networks, informing cyber security officers of most likely and recent incursions from both amateur and sophisticated threat actors. Cyber Threat Intelligence (CTI) is about contextual information allowing stakeholders to make decisions as cyber-attacks progress. A known drawback in security analysis is to use simple Indicators of Attacks (IoA) or similarly Indicators of Compromise (IoC) that are easily circumvented by adversaries as their attacks increase in sophistication, determination, and creativity when corrupting systems and data. CTI bridges this technical gap by working not only with IoA/IoC to sustain their analysis, but also incorporating Tactics, Techniques, and Procedures (TTP), i.e., the 'why' and 'how' cyber-attacks could be potentially perpetrated. Central to CTI is *timeliness* inspecting how to adequately respond to impending malicious incursions to quickly identify and confirm criminal activities before they propagate over networks.

This paper will expand on the prospects of employing social media-based data for analysis and practical use in cyber security employing SNA. We shall present two case studies, one for Twitter and another one for Reddit, showcasing features and limitations. Our idea centres on working with validated users/accounts that consistently disseminate accurate and timely information to their audiences. The approach thus collects and enumerates such accounts, devise a *trustability metric* so it establishes trust in the community, and finally, it processes and presents this timely information to subscribers of our intelligence service.

The paper is organised as follows. Section 2 will outline recent approaches in CTI and social media datasets whereas Section 3 discusses our approach on how to extract meaning information. We proceed to Section 4 on a case study for Twitter and Section 5 with a Reddit crawler. We end our work in Section 6 with final considerations and ideas moving forward.

2 Related work

The intelligence community considers crowd-based approaches to help gathering and analysing information from different sources that are public or generated by devices. There is a distinction among types for *intelligence crowdsourcing* [18], namely Open-Source (OSINT), Human (HUMINT), Signals (SIGINT) and Measurement and Signature (MASINT) Intelligence, among a few other gathering disciplines. In the context of this work we shall focus on OSINT [22], due to focusing on social media outlets. It has attracted scientific attention due to its relevance on gathering indications that could lead to both cyber-attacks and mitigation efforts trusting in the collective and collaborative nature of public social networks [25,24,17].

As mentioned, TTPs focus on 'why' and 'how' sophisticated threat actors traverse networks and systems for vulnerabilities. Organisations have proposed

different frameworks for modelling TTP throughout the years, such has the Adversarial Tactics, Techniques & Common Knowledge (ATT&CK®) framework[1] [19] compiles common TTPs employed by malicious adversaries to help establishing mitigations and controls to thwart their advances.

Mavroeidis and Bromander [11] discussed taxonomies and standards for CTI. MITRE has developed interesting CTI based information under the ATT&CK framework [19] where Xiong et al. [23] have devised threat modelling for advanced security analysis. The framework is sustained by the Structured Threat Information eXpression (STIX™) format, which is a standard for creating CTI models and sharing, employed by Czekster et al. [3] in a general web-system for describing cyber-attacks generating as output STIX format models.

In terms of quantitative metrics explored in social media, we highlight the work of Gräve (2019) [5] that tried to measure the impact of social media communication. Peters et al. (2013) [14] discussed social media based metrics and the impact on subsequent evaluations. These researches point out the need for better understanding and working with metrics to enhance assessments and analysis.

CTI is used across different application domains and it is particularly useful for capturing so called on-line chatter that could substantiate security analysis [16]. In the context of smart devices, it could use STIX for leveraging CTI taking into account their functionality and interaction [2]. Specifically towards the use of Twitter, we mention Riebe et al. (2021) [15] tool called `Cysecalert` for generating alerts using OSINT, `Twitterosint` [6], a tool for automated collection, analysis, and visualisation of intelligence, and `Sonar` [10], a system for detection of cyber security events using Twitter streams. We highlight `Iocminer` [12], a tool for inspecting IoC in Twitter that combines automatic graphs and text mining for CTI. The authors explore using regular expressions to extract meaningful instances from twits meriting further inspection.

3 Extracting information from social media

As mentioned, social media networks may produce a massive amount of data by numerous individuals and organisations. It is a challenge to sift through all the data and uncover relevant information concerning specific problems. Our focus here is on OSINT and actionable intelligence gathering, so we must take steps to devise an approach that quickly sorts out results and presents to users so they make timely decisions. We want to avoid situations where the tool keeps processing and communicating instead of presenting users with relevant information, so our framework should be equipped with features to inspect performance and data retrieval as it is executing.

We detail next the four-step process and explain our guided user-oriented tool interaction altogether:

[1] MITRE's ATT&CK Framework: `https://attack.mitre.org/`.

1. Graphical User Interface (GUI): it presents users with a simple to use, friendly interface where they assign preferences, filters, and advanced options that will focus responses to meet their expectations.
2. Identification, mapping phase: to identify validated and trustworthy accounts or data sources that disseminate useful and timely information to their respective communities.
3. Tool internal execution: internally, we will use specific Application Programming Interface (API), depending on the social media, to subscribe to any events that are published by these selected accounts. It retrieves, processes and analyses data intersecting with filters put forth by users of our tool.
4. Presentation: it visually represents all gathered intelligence to the user, ordered by time, labelling it with the corresponding social media outlet to enact faster decision making and enabling quick thwarting cyber-attacks or malicious incursions before they propagate.
   – After presenting to our user base, it collects interactions with the information (such as likes/dislikes, upvotes/downvotes, etc.) so it updates top accounts for relevancy, altering the order on which to show results.

Figure 1 showcases a general approach with the necessary steps to conduct a timely analysis for intelligence gathering over different social networks outlets.

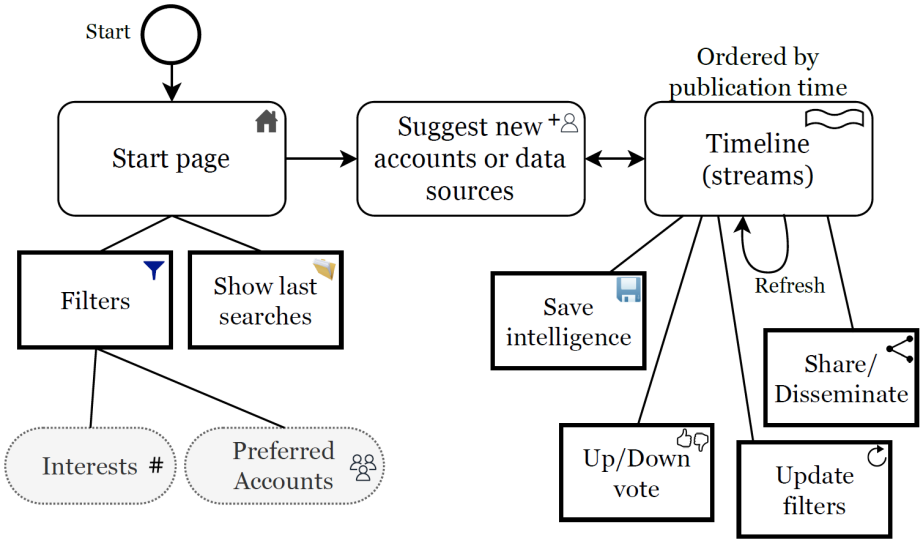

Fig. 1. General process of using social media outlets to increase analysis efforts.

A key element of our proposition is to choose reliable, trustworthy, and validated social media accounts or certified data sources. These users will vary depending on selected interests and filters as set out by our user base. Our approach will present a pre-selected number of manually inspected accounts or

data sources pruned for relevance and useful content. Our users may signal different accounts to employ, and we shall take those into account and validation to present to the rest of the user base case it meets our relevance criteria.

3.1 Working on a *trustability* metric

Given the multitude of accounts and social media networks, one interesting problem is to determine which venues to trust, which are accurate (over time), separating these selected accounts for adding to our tool's subscription mechanism. We are interested in computing a numerical value per account that indicates its level of so called *trustability*, i.e., how trustworthy such account is to be taken into consideration by our future analysis. This metric should allow new accounts to be at least visible, so as to "earn one's trust" and thus enter the selected accounts list. For these accounts, we will not consider the date on which they started operating.

There are key quantitative indices to inspect to build our trustability metric, for instance, the number of followers, number of views, quantity of reposts/share/re-tweets (each platform has its own naming), number of comments (if any) or user interactions, and a network-based property that considers how the accounts are connected altogether.

Note that there exist third-party (often paid) services that attempts to inflate the number of followers artificially has been a noticeable issue since the very beginning of social networks as they are a quality measure of content creators and their outreach over communities. Known problems plaguing social networks in general are the number of fake users, i.e., (non-validated) accounts that promotes disinformation and skew analysis efforts. This is why developing and computing a metric poses significant challenges for researchers, and the number might not represent the reality.

The idea behind many social media instances is to keep users *locked in* the platform. They do that by showing content indefinitely (e.g., infinite scrolling) as the algorithms are familiar with the things users enjoy consuming. The platforms have customised the algorithms to a point where the user simply can't leave, however, this has egregious consequences to young populace that displays symptoms similar to addiction [9].

Table 1 shows key characteristics of known social media outlets and the kind of data algorithms work to present new content to their user base. Column *Content track* indicates how one interacts with the platform, feeding it with responses in agreement, disagreement, or 'mood'. For example, "heart" (as outlined with ♡) in Instagram corresponds to a positive response to the content.

A lot of social media outlets focus on providing users with a good, fun online experience, so they will not track (or show) any numerical measurement indicating any kind of stress like the number of dislikes or how many people have *unfollowed* an account over time. It is possible to infer those values, however, sometimes they are simple not available for any kind of analysis.

Another substantial problem is the fact that many users simply do not interact with social platforms in any way (campers/lurkers/etc.), however, this is

Table 1. Characteristics of engagement across different social media outlets.

Social Media outlet	***Linking* (naming)**	**Dissemination**	**Content track**	**Additional tracking**
Facebook	Friend	Share, mark	Like, love, care, haha, wow, sad, angry	Comments
Instagram	Follow	Share	♡	Comments
TikTok	Follow	Share	♡	Comments
Threads	Follow	Repost, share, mention	♡	Replies, comments
Twitter/𝕏	Follow	Retweet, mention	♡	Comments
YouTube	Subscribe	Share	I like this 👍, I dislike this 👎	Views, comments

detrimental to the social media algorithms as they should rely only on viewing content instead of actively engaging through like/dislike (or similar interaction) so it performs better (users understand that they should interact so similar content is shown in the future).

Table 2 outlines the variables we are looking for deriving a formula to compute the proposed trustability metric.

Table 2. Variables we considered to derive our trustability metric.

Characteristic	**Symbol**	**Comments**
Register date	D	Scale: days ago
Validated/verified account?	V	Possible values: 0 or 1
Date of first publication	Df	Scale: days
Last login or last seen	L	Scale: days ago
Number of followers	F	Numerical value
Total content	Tc	All produced content
Content Creation Rate	CCR	$CCR = \frac{Tc-Df+L}{D}$
Views	Vi	Numerical value
Virtual reward	R	Value ranging from 0 to 1
Total number of comments	Co	Total number of comments

We propose a Trustability Metric TM as shown in Equation 1, concerning individual social media accounts, thus follows:

$$TM = \left(F \times \frac{\overbrace{Tc - Df + L}^{CCR}}{D} \times \frac{\overbrace{1}^{engagement}}{Co \times Vi \times R} \right) \times (V + 1) \qquad (1)$$

Our formula for computing TM takes into account the user engagement towards the account's produced content over the period on which it is active, yielding a measure of trust that could be used to decide whether to follow it for intelligence gathering. CCR takes all produced content Tc and discounts when the user first posted anything plus the last time they logged in over the day of registration. This is an attempt to determine engagement and avoid passive or camper users, ie, those that just consume content without ever generating it.

One characteristic that does exist but we have not considered is the type of account (business or user) as renown organisations when publishing intelligence pieces could add more trustworthiness to the analysis. This will be the subject of future research. Table 3.1 tests the metric outlining synthetic accounts populated with quantitative random data.

Table 3. Evaluation of TM for randomly generated values in social media accounts.

Synthetic Account†	**D**	**V**	**Df**	**L**	**F**	**Tc**	**CCR**	**Vi**	**R**	**Co**	**TM**	**TM'**‡
`@abc123` Twitter	88	1	51	1	23,450	405	4.034	233,892	0.1	4,334	0.0004	2.729
`@hugh3` YouTube	5	0	5	0	100	2	-0.600	56	0.05	12	-1.7857	-0.252
`@kikin3tw` Instagram	102	1	10	2	1,302	25	0.167	8,391	0.1	1,221	0.0001	3.373
`@baba00b` Twitter	1,011	1	1,000	10	123,555	1,010	0.020	90,000	0.05	1,234	0.0002	3.055
`@ju_b1bopr` YouTube	123	0	23	20	111	5	0.016	590	0.1	10	0.0030	2.514

†These accounts might exist by pure chance.
‡Formula: $TM' = -log10(|TM|)$, higher is best.

It is noticeable that TM prioritises quality and interaction over quantity. It privileges the amount of content over engagement and user interaction through comments, views, and so called virtual rewards, i.e., a number between 0 and 1 that indicates how many users actively clicked on any button representing agreement or disagreement (like/dislike, heart/*un-heart*, etc.).

Limitations: The metric is not authoritative, and act as a suggestion to map and use a quantitative measure for trust, and further research is required. Security and intelligence analysis must consider other approaches when choosing their sources to work. The idea here is to use publicly available measures to compute this index that are most likely available through API calls or straightforward crawling outputs (if permitted) within each platform.

The trustability metric is crucial to determine the quality of the sources under intelligence gathering that security analysts should consider and prioritise over other information venues. The mechanism we describe herein subscribes to key accounts and follow them in their respective platforms, having access to their stream and feeds for tracking worthy intelligence pieces altogether.

One alternative that could be used by analysts is to devise a bespoke index by adjusting TM' to a factor that address the community's support to a given account. Originally, the proposed trustability metric uses only quantitative data available from accounts and user interaction; using this factor might add to the index how trustworthy the community sees the accounts, which might foster better analysis throughout.

4 Case study on Twitter: the CyberTweets tool

In this paper we are interested in designing and implementing tool that collects and organises OSINT-based data from social media. We have developed a tool called CyberTweets for extracting Twitter based data and convey it to end-users. It is a web-based tool which connects to Twitter API v2 to get cyber security related tweets and analyses them further to display the data to the user in a user-friendly way. CyberTweets website offers not only nice interface but also many features which could be useful for cyber analysts or IT professionals.

4.1 Operational details and architecture

CyberTweets utilises natural language processing (NLP) techniques to detect useful information for threat intelligence. In terms of its architecture, it was built using React (18.2.0) for the front-end, NodeJS (16.14.2) and Express for the server side and MongoDB for the database. Figure 2 illustrates CyberTweets' basic architecture and interaction with other elements from the system.

We have used other auxiliary libraries and APIs to build up the application namely `CORS` (for cross-origin HTTP requests), `Dotenv` (to manage environment variables), `Mongoose` (connections to MongoDB), `Express-async-handler` (to simplify error handling), `Bcryptjs` (for basic encryption features), `Jsonwebtoken` (session management), `Axios` (library for making HTTP requests), `Node Cron` (for scheduling), `Node NLP` (analysing data for NLP), `MUI` (interfacing library), and Styled components (dynamic rendering of webpages). The tool was hosted at Amazon Web Services (AWS) for simplifying Continuous Integration and Continuous Deployment (CI/CD), among other benefits for the basic user account

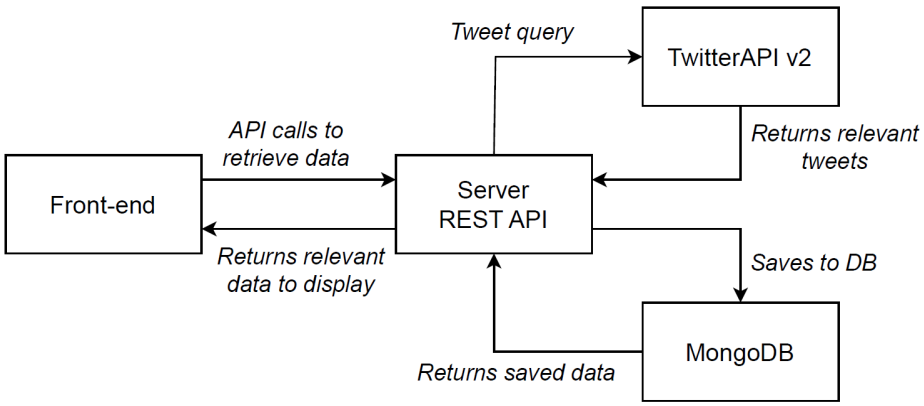

Fig. 2. Overview of CyberTweets' operational details and architecture.

included in the offer. The application uses the Twitter API v2 and data matching pre-set requirements, such as domain, entity fields, hashtags, account ID, and the tweet ID. We feed this information to the MongoDB database as soon as the request processes finishes.

4.2 Feature set and basic screens

The tool allows users to filter tweets by date range, hashtags, search text, and named entities identified using the Node-NLP library. Additionally, users can save and download the text data from tweets. The project aim is to research and develop a useful tool which allows OSINT data gathering from Twitter and offers ways to analyse it further for threat identification, potentially helping many cyber security experts in their daily activities. Figure 3 highlights the opening screen for CyberTweets application, assuming the user has logged in and has had already chosen interesting topics and accounts to follow.

The tool shows current filters (logged users), list of applied filters and additional options (by date, by source, by entities, e.g., 'malware', 'hacker groups', or 'organisations'), plus a "Search" feature to look out pieces by providing text snippets according to user aims. Figure 4 shows how to set user preferences.

In this current version, the application is manually following trusted accounts instead of using the trustability metric. In the near future, we envision to integrate and test using the metric to determine and suggest users other interesting accounts they should integrate into their analysis.

4.3 Analysis and end-user evaluation

CyberTweets website has been evaluated by a few test users which have IT expertise to measure its potential use. We have conducted a basic usability testing over a reduced set of three users, using a qualitative approach to reason on less

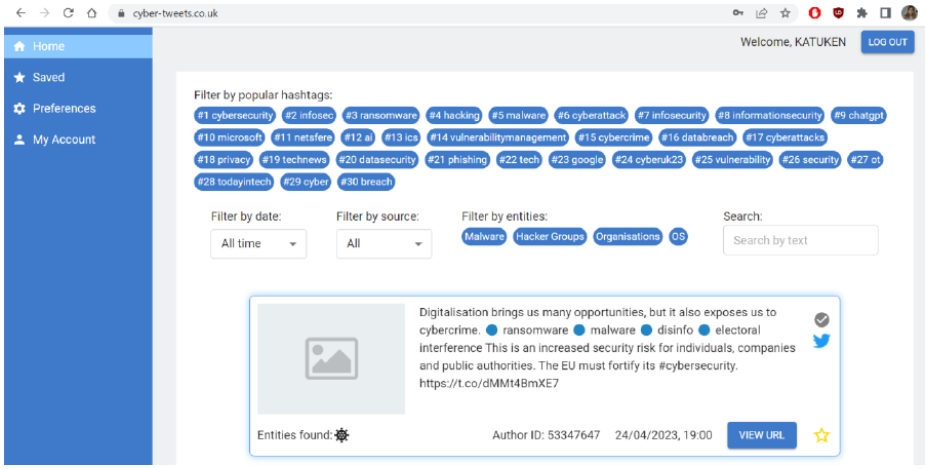

Fig. 3. Welcome screen for the CyberTweets application and basic feature set.

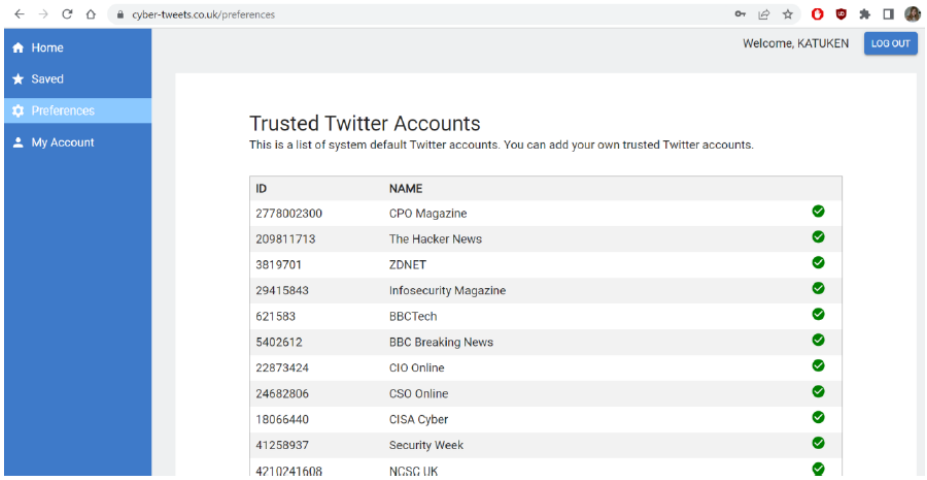

Fig. 4. Preferences for users, showing accounts that it is currently following.

usable parts of the system and where improvements could be tackled for better understanding how to approach intelligence tasks. We have arranged interviews with users, where they were exposed to the tool, performed tasks, and answered a questionnaire following the process, so we understand their reasoning and overall perception whilst using the tool. We highlight a list of nine tasks we asked users to conduct, and the time they took to comply (following pattern `Task#n Description [average time, in seconds]`):

Task#1: Find tweets older than three days [17.3]
Task#2: Save a tweet [7.0]

Task#3: Find a tweet which has a hashtag you are interested in [15.3]
Task#4: Find a tweet which has an entity of potential interest to security [7.6]
Task#5: Change the list of trusted accounts (default accounts are added, add a new account with an ID number as `123456` and the name as `Test`) [36.3]
Task#6: Access your saved tweets [1.6]
Task#7: Change your name (note that your current name is set to `TestUser`, please change it to `UserTesting`) [8.3]
Task#8: Change your password (dummy password is `Testing123`, please change it to `Testing567`) [15.0]
Task#9: Find some specific information in the tweet text [10.3]

The most difficult task they performed was Task#5, when changing the list of trusted accounts. This has suggested us the need to improve ways when presenting users with simpler mechanisms for better understanding the concept of trusted accounts and how they could easily access it or present it in a seamless fashion. We present questions used in the interview with users:

1. What is your overall perception of the tool?
2. In your opinion, does it have the potential to attract other users?
3. Do you know any tool that is related to this one? If yes, what it is called?
4. Are icons easy to understand? For example, the tick, star etc.?
5. Was the website easy to use and navigate?
6. Were there any website features that you particularly liked or disliked?
7. Did you find the filters (date, hashtags, entities, search) helpful in narrowing down the tweets you wanted to see?
8. Did you find the information provided by the CyberTweets website useful?
9. Was there any other functionality you would have liked to see?
10. Do you have any recommendations to improve the tool?

The feedback received was generally positive where users have provided valuable insights for improving the tool in future versions. Some users found the tasks very easy to do but some of them struggled, either not understanding it or it just was a user's preference to perform it in a different way. When asked to find a tweet by a popular hashtag one user used a search field and the other checked the tweet text manually. Another interesting observation was that when asked to add a new trusted twitter account none of the users clicked on `Preferences` tab first, most clicked on `My Account`. This indicates that perhaps the naming of Preferences should be changed.

5 Case study on Reddit: webcrawler to STIX bundle

We proceed working with Reddit, a popular social news aggregator and discussion board. We have created an account for crawling data from specific directories (called 'subreddits' in the platform) namely cyber security related discussion fora. Having an account and authenticating in the platform will extend the limitations when retrieving data. The crawler is a Python script integrated with The Python Reddit API Wrapper (PRAW)[2] – see Appendix A for code excerpts.

[2] Link: `https://github.com/praw-dev/praw`.

After deciding which data for CTI, we will convert the set of strings into STIX [7,1] format and generate a STIX model that could be consumed by the cyberaCTIve tool [3,4]. This tool is publicly available[3] and allow users to directly insert CTI streams acting as a front-end to the STIX format. It presents users with available parameters (mandatory and optional) and offers interesting features such as attack timeline analysis and sharing STIX models (it dumps the model representation for storage and sharing into other tools).

Given the reduced set of attributed that can be retrieved by PRAW dictated by the platform's limits, we will need to modify the way we computed the trustability metric to accommodate the relevance of submissions into cyber threat analysis. For instance, the PRAW's object `Redditor` will only provide the user's name and `comment_karma` (this attribute is somewhat related to reputation) and the date they first logged in (under `created_utc` attribute) whilst for the object `Submission` we had available its `title`, `score`, `upvote_ratio` (i.e., the percentage of upvotes from all votes on the submission), `num_comments`, and `date_created` (unless detailed, these attributes are self-explanatory).

Figure 5 outlines our general process for gathering intelligence out of social media outlets, particularly Reddit.

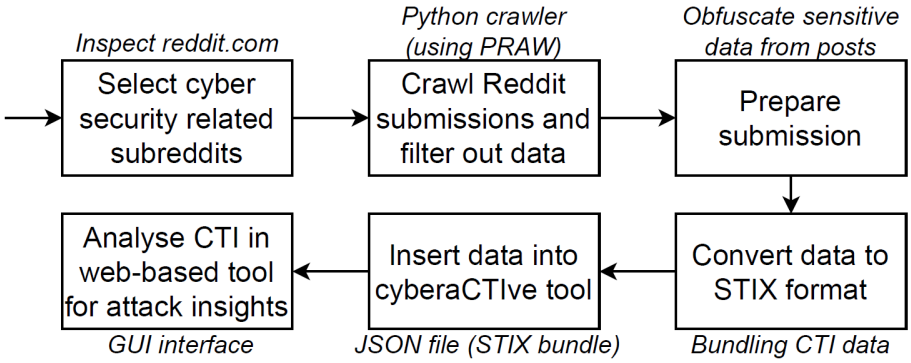

Fig. 5. Converting posts from Reddit to actionable CTI.

Reddit imposes limits to submission retrieval for non-authenticated and authenticated users/crawlers. As of Sep/2023 the documentation states that *"Any single reddit listing will display at most 1000 items."* and that *"Apps that make less than 100 queries per minute per OAuth ID can still use the free tier"*, i.e., one must resort to filter out submissions by `hot()`, `new()`, or `top()` (employing limits to the functions such as `subreddit.hot(limit=50)`) as data crawlers must comply with this to get any meaningful results.

Reddit offers a host of subreddits related to cyber security. One notices an active community involving non-technical, developers, designers, software quality

[3] Link: `https://cyberactive.performanceware.com.br/`.

and pentesters where cyber security is the focal point. Next, we present a list of subreddits that we could potentially employ to extract meaningful CTI. We note that this is not comprehensive, i.e., there will be many other subreddits within the platform that discuss the topic of cyber security. All links will work adding to the browser the prefix `https://www.reddit.com/` followed by the subreddit's link as mentioned below. We also state the number of members within brackets ([N]) – as of September 2023 – and that can be used to prioritise certain subreddits for CTI as well. Unless stated otherwise, these subreddits are public.

- `r/BadApps` [2.7k]: all information about malware (focus on smart phone apps and Google's Play Store).
- `r/blackhat` [83.5k]: vulnerabilities and exploitation.
- `r/blueteamsec` [38.1k]: focuses on technical intelligence for blue teams.
- `r/computerforensics` [59.7k]: dedicated to recovery and investigation of cyber related materials.
- `r/ComputerSecurity` [31.6k]: provides curated list of links to IT security.
- `r/cyber` (private): covers the geopolitical and corporate CTI.
- `r/cybersecurity` [564k]: general discussion about cyber security, careers, developing threats.
- `r/ethicalhacking` [27.7k]: interest in computer hacking (ethically).
- `r/ExploitDev` [11.2k]: discusses software vulnerabilities and exploits.
- `r/fulldisclosure` [2.2k]: relates information about breaches, exploits, data leaks and vulnerabilities.
- `r/hackersec` (private): offers technical guides to interact and share information about cyber security.
- `r/hacking` [2.6m]: dedicated to computer hackers and hacking in general.
- `r/Information_Security` [22.3k]: publishes information about security news and analysis.
- `r/Malware` [70.9k]: discussions on malware reports and related information.
- `r/pwned` (private): news about recent exploits and breaches, leaked data.
- `r/redteamsec` [25.1k]: converges red and blue teams to discuss malware, tradecraft and reverse engineering.
- `r/reverseengineering` [135k]: discusses topics related to reverse engineering software, breaking it apart and understanding it.
- `r/threatintel` [3.7k]: discusses threats and sharing information across stakeholders to thwart the advance of attacks.
- `r/websecurity` [6.2k]: covers links and discussions about development and security of websites, aggregating owners, developers and pentesters.

Out of this list of subreddits, we point out that `r/hacking` has about 2.6 million members so one must factor in the sheer amount of noise that such large community creates over time. On the other hand, it may act as a valuable source for useful CTI for understanding attacks. The main cyber security subreddit (`r/cybersecurity`) with 564 thousand users stands out as a valid alternative that might yield an interesting amount of data pieces for CTI.

5.1 Retrieving data from selected subreddits into cyberaCTIve

For this work we have decided to retrieve only hot and new first 50 submissions and compute the trustability metric to determine which posts merited to be further analysed. Out of those, we extracted the submission's title and looked for interesting (with respect to CTI) descriptions worth creating STIX objects capable of feeding the cyberaCTIve on-line tool.

If we do not limit the list of subreddits to inspect, the script will produce an output that has a lot of noise to process. We have chosen to look into the following ones: `"r/cybersecurity"`, `"r/ethicalhacking"`, `"r/malware"`, `"r/threatintel"` and `"r/blueteamsec"`. In light of this fact we have chosen cyber security related subreddits to search for strings having the following:

```
regexes = [ "attack.*", "attacker.*", "cybersecurity.*", "CVE.*",
            "NVD.*", "security.*", "malware.*", "threat actor.*",
            "threat.*", "MS-Windows.*", "Android.*", "MacOS.*",
            "iOS.*", "GNU-Linux.*", "Linux.*", "vulnerability*" ]
```

where `regexes` is a Python data structure to withhold all our strings for pattern matching using regular expressions. This is a clear limitation of our approach as one must fine-tune it from time to time to capture other ways redditors are creating CTI related content.

For this task here, we are interested in running a periodic Python script that will traverse new posts, extract relevant CTI data, and convert to standardised formats (namely STIX) for later analysis. We are tracking each individual posts' identifiers to avoid creating repeated entries in JSON files.

Each post shall map into distinct STIX objects as follows. As STIX Domain Objects (SDO) we could add data into `infrastructure`, `malware`, `malware analysis`, `note`, `observed-data`, `opinion`, `report`, `threat-actor`, `tool`, and `vulnerability` and STIX Relationship Objects (SRO) such as the `relationship` type. For STIX Cyber-observable Objects (SCO), we could potentially use `file`, `ipv4-addr/ipv6-addr`, `network-traffic`, `software`, `url`, and `user-account`, to name a few. The mapping will depend on our ability to understand the post and assign meaning to it, in the hope of extracting useful and actionable CTI.

We present Algorithm 1 (named `Reddit2CTI_Converter`) that crawls Reddit for CTI-related content as explained next.

From the list of subreddits and the regular expressions to perform the pattern matching, the script will traverse the list of posts per subreddit and select likely CTI to populate cyberaCTIve tool. The function `crawl_reddit` does the data extraction within Reddit whereas the `assign_type` will determine which STIX object to assign to the post. Function `compute_trustability` will take the new entry and the discussion, extract the current Redditor and determine the trustability metric as explained in Section 3.1. We have created a function called `acceptable(tm)` that evaluates the trustability metric and determines whether the piece of data merits further analysis, i.e., insertion into the CTI tool. Finally, the function `convert` will transform the post into a JSON file and

Algorithm 1 Reddit2CTI_Converter

subreddits ← list of selected (relevant) to cyber security Reddit subreddits
regexes ← list of strings related to CTI tracking
for all *subreddit* : *subreddits* **do** ▷ traverse the list of selected subreddits
  discussions ← *get_discussions*(*subreddit*, *regexes*) ▷ match discussions
  for all *discussion* : *discussions* **do** ▷ process all discussions
    entry_new ← *crawl_reddit*(*subreddit*, *regexes*) ▷ match discussions
    tm ← *compute_trustability*(*entry_new*, *discussion*)
    if *acceptable*(*tm*) & *unique*(*entry_new*) **then** ▷ If this is a new item
      type ← *assign_type*(*entry_new*) ▷ set STIX objects: SDO, SRO, SCO
      entry_json ← *convert*(*entry_new*, *type*) ▷ returns JSON file for entry
      insert(*entry_json*) ▷ Insert JSON file into cyberaCTIve
    end if
  end for
end for

function `insert` that could feed it into the cyberaCTIve tool[4]. The function will also create a new STIX model, assign temporary names and STIX objects parameters depending on each entry.

We have implemented the algorithm explained in previous section as a Python script integrated with PRAW. Next, we show Python excerpts to showcase our approach:

```
import praw # if not present, run 'pip install praw'

def main():
    reddit = praw.Reddit(
        client_id=client_id,         # given by Reddit
        client_secret=client_secret, # given by Reddit
        password=my_pass,            # chosen by user
        username=my_user,            # chosen by user
        user_agent=my_useragent,     # create a mock user agent
        check_for_async=False        # prevent warning
    )
    my_subreddits = [
        "cybersecurity",
        "ethicalhacking",
        "malware",
        "threatintel",
        "blueteamsec"
    ]
    regexes = [
        "attack.*", "attacker.*", "cybersecurity.*",
        "security.*", "malware.*", "threat.*",
```

[4] Note that this feature is not yet implemented in the Python crawler described herein.

```
        "threat actor.*", "CVE.*", "NVD.*", "vulnerability*",
        "MS-Windows.*", "Android.*", "MacOS.*",
        "iOS.*", "GNU-Linux.*", "Linux.*"
    ]
    combined = "(" + ")|(".join(regexes) + ")"
    for mysubreddit in my_subreddits:
        subreddit = reddit.subreddit(mysubreddit)
        i=0
        print("Subreddit: " + mysubreddit)
        #pick one method for retrieving posts out of this options:
        #submissions = subreddit.stream.submissions()
        #submissions = subreddit.new()
        #submissions = reddit.subreddit("all").search(mysubreddit,
            sort="hot", syntax=None, limit=LIMIT)
                # 'sort': "relevance", "hot", "top", "new", or "comments".
        submissions = subreddit.hot(limit=LIMIT)
        for submission in submissions:
          if (hasattr(submission.author, 'name') and   # checks valid user
            user_exists(reddit, submission.author) and # checks valid user
            submission.stickied == False and           # checks only 'unpinned'
                                                       #  ('unstickied') items
            re.search(combined, submission.title)):
            i=i+1
            process_submission(i, submission)
    file1.close()

def user_exists(reddit, name):
    try:
        if hasattr(reddit.redditor(name), 'id'):
            reddit.redditor(name).id
    except NotFound:
        return False
    return True

def process_submission(i, submission):
    file1.write("\n\nProcessing submission [" + submission.permalink + "]")
    if hasattr(submission.author, 'created_utc'):
        date = arrow.get(submission.author.created_utc).to('local').humanize()
    file1.write("---\nRedditor: " + submission.author.name +
          " (karma: " + (str(submission.author.comment_karma)
              if hasattr(submission.author, 'comment_karma') else "") +  ")")
    file1.write("\Title: " + str(i) + ": " + submission.title +
          ", score:" + str(submission.score) +
          ", upvote_ratio: " + str(submission.upvote_ratio) +
          ", num comments: " + str(submission.num_comments)
```

```
        "")

if __name__ == "__main__":
    main()
```

The idea is to crawl Reddit for specific data concerning cyber security and process it in our end to filter out unwanted results and prioritise elements we deem important.

Next, we show one instance of our *proof-of-concept script* (subject to further validation in the near future) return to showcase how the conversion to STIX could work (note that each redditor's name was obfuscated) from the subreddit `r/Malware`.

```
Permalink: /r/Malware/comments/15136a4/malware_delivery_
           via_microsoft_teams_law_firms/
Redditor: dkal33ks2 (karma: 8)
Title: Malware delivery via Microsoft Teams, law firms under
       cyberattack, CVSS 4.0 is out
Statistics: Score= 6, upvote_ratio= 0.75, num comments= 3
```

In terms of converting these data intelligence, it consists on creating a STIX model with a pre-selected name and create a bundle in JSON format. For the first item, one might come up with the following ***excerpt*** of STIX model:

```
{ "type": "bundle",
  "id": "bundle--ba83f63f-72ff-4f49-9fca-616f2d29e13c",
  "objects": [
  {
    "type": "infrastructure",
    "id": "infrastructure--ed75baaf-54eb-472c-b68a-4cdd0731fa13",
    "name": "Company PC on third floor"
  }, {
    "type": "malware",
    "name": "reddit-delivery-via-ms-teams",
    "malware_types": [ "backdoor" ],
    "implementation_languages": [ "c++" ]
  }, {
    "type": "relationship",
    "id": "relationship--0aaab1f3-e296-4860-8980-238d3d2abf10",
    "source_ref": "infrastructure--ed75baaf-54eb-472c-b68a-4cdd0731fa13",
    "target_ref": "malware--30494a3a-899f-451e-ad10-264eb7873c53"
  }, { ... } ] }  # as the STIX model will have more entries
```

The difficult part is not the conversion to STIX itself, but to have interesting intelligence pieces as input for the implementation. Developers working on a subsequent tool should focus on the Reddit crawler, i.e., determining what is and what it is not considered intelligence, and then calling a conversion function.

5.2 Discussion

Extracting CTI out of Reddit is not trivial given the multiple ways the community is creating messages and discussions across varied topics. We stress that a lot of internal details from this tool were omitted here, specially when converting posts into the STIX format and general modelling.

PRAW is a satisfactory tool with simple API to access Reddit, however, creating working Python scripts requires a lot of tweaking and checks to make it work. The way of sorting posts should employ programming after the crawling process altogether, i.e., the API is limited, perhaps to prevent abusing the platform through scripting. Sometimes data does not exist (e.g., has been removed by the platform, however, it still returns the post by the API), the user has been removed, or the post has specific attributes (stick or flagged) by moderators, among other issues.

6 Conclusion

Present work has two main contributions, firstly, it has shown how to employ social networks data sources as complementary intelligence gathering (secondary sources), and secondly, case studies in Twitter and Reddit, showcasing interesting examples of how to build a tiered system for handling intelligence. We have highlighted the main issues and concerns when implementing such systems, discussing a design framework that meets social networks elements and quantitative data that can be used to compute a trustability metric.

There has been interesting discussions, challenges, and open research ideas on how to best approach OSINT to explore its use in daily cyber security mitigation efforts for both end-users and organisations [13]. As future work for CyberTweets we will integrate user feedback and expand the feature set with other suggestions, coupling the system with automated account subscription emerging in social networks, adding more filtering options, and collating data from multiple platforms altogether.

For Reddit-based CTI retrieval there is the need to fine-tune the search strings to match more meaningful data items out of the platform as well as increasing the number of subreddits to inspect. One notices that expressing the most relevant way to extract intelligence from these social networks is not trivial and merits further research. Ultimately, CTI is about achieving a balance among employing high-quality sources, timeliness, and quickly separating signals from noise for quick analysis and response. As a final note, we are conscious of current efforts on employing and combining ideas in Artificial Intelligence/Machine Learning (AI/ML) in OSINT-based research, exploring natural language processing and integrating with other information systems in organisations.

References

1. Barnum, S.: Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX). Mitre Corporation **11**, 1–22 (2012)

2. Czekster, R.M.: Leveraging Cyber Threat Intelligence in Smart Devices. In: Information Security and Privacy in Smart Devices: Tools, Methods, and Applications, pp. 71–95. IGI Global (2023)
3. Czekster, R.M., Metere, R., Morisset, C.: cyberaCTIve: a STIX-based Tool for Cyber Threat Intelligence in Complex Models (2022). https://doi.org/10.48550/ARXIV.2204.03676
4. Czekster, R.M., Metere, R., Morisset, C.: Incorporating Cyber Threat Intelligence into Complex Cyber-Physical Systems: A STIX Model for Active Buildings. Applied Sciences **12**(10), 5005 (2022)
5. Gräve, J.F.: What kpis are key? evaluating performance metrics for social media influencers. Social Media+ Society **5**(3), 2056305119865475 (2019)
6. Hoppa, M.A., Debb, S.M., Hsieh, G., KCa, B.: Twitterosint: Automated open source intelligence collection, analysis & visualization tool. Annual Review of Cybertherapy And Telemedicine 2019 **121** (2019)
7. Jordan, B., Piazza, R., Darley, T.: STIX Version 2.1. OASIS Standard, Available on https://docs.oasis-open.org/cti/stix/v2.1/os/stix-v2.1-os.html, Accessed: 01/Sep/2023. (2021)
8. Kong, X., Shi, Y., Yu, S., Liu, J., Xia, F.: Academic social networks: Modeling, analysis, mining and applications. Journal of Network and Computer Applications **132**, 86–103 (2019)
9. Kuss, D.J., Griffiths, M.D.: Social networking sites and addiction: Ten lessons learned. International journal of environmental research and public health **14**(3), 311 (2017)
10. Le Sceller, Q., Karbab, E.B., Debbabi, M., Iqbal, F.: Sonar: Automatic detection of cyber security events over the twitter stream. In: Proceedings of the 12th International Conference on Availability, Reliability and Security. pp. 1–11 (2017)
11. Mavroeidis, V., Bromander, S.: Cyber threat intelligence model: an evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence. In: 2017 European Intelligence and Security Informatics Conference (EISIC). pp. 91–98. IEEE (2017)
12. Niakanlahiji, A., Safarnejad, L., Harper, R., Chu, B.T.: Iocminer: Automatic extraction of indicators of compromise from twitter. In: 2019 IEEE International Conference on Big Data (Big Data). pp. 4747–4754. IEEE (2019)
13. Pastor-Galindo, J., Nespoli, P., Mármol, F.G., Pérez, G.M.: The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends. IEEE Access **8**, 10282–10304 (2020)
14. Peters, K., Chen, Y., Kaplan, A.M., Ognibeni, B., Pauwels, K.: Social media metrics—a framework and guidelines for managing social media. Journal of interactive marketing **27**(4), 281–298 (2013)
15. Riebe, T., Wirth, T., Bayer, M., Kühn, P., Kaufhold, M.A., Knauthe, V., Guthe, S., Reuter, C.: Cysecalert: An alert generation system for cyber security events using open source intelligence data. In: International Conference on Information and Communications Security. pp. 429–446. Springer (2021)
16. Sapienza, A., Ernala, S.K., Bessi, A., Lerman, K., Ferrara, E.: Discover: Mining online chatter for emerging cyber threats. In: Companion Proceedings of the The Web Conference 2018. pp. 983–990 (2018)
17. Soomro, T.R., Hussain, M.: Social Media-Related Cybercrimes and Techniques for Their Prevention. Appl. Comput. Syst. **24**(1), 9–17 (2019)
18. Stottlemyre, S.A.: HUMINT, OSINT, or something new? Defining crowdsourced intelligence. International Journal of Intelligence and CounterIntelligence **28**(3), 578–589 (2015)

19. Strom, B.E., Applebaum, A., Miller, D.P., Nickels, K.C., Pennington, A.G., Thomas, C.B.: MITRE ATT&CK: Design and philosophy. Technical report (2018)
20. Tabassum, S., Pereira, F.S., Fernandes, S., Gama, J.: Social network analysis: An overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **8**(5), e1256 (2018)
21. Wasserman, S., Faust, K.: Social network analysis: Methods and applications. Cambridge University Press (1994)
22. Williams, H.J., Blum, I.: Defining second generation open source intelligence (OSINT) for the defense enterprise. Rand Corporation Santa Monica (2018)
23. Xiong, W., Legrand, E., Åberg, O., Lagerström, R.: Cyber security threat modeling based on the MITRE Enterprise ATT&CK Matrix. Software and Systems Modeling **21**(1), 157–177 (2022)
24. Yadav, A., Kumar, A., Singh, V.: Open-source intelligence: a comprehensive review of the current state, applications and future perspectives in cyber security. Artificial Intelligence Review pp. 1–32 (2023)
25. Yeboah-Ofori, A., Brimicombe, A.: Cyber intelligence and OSINT: Developing mitigation techniques against cybercrime threats on social media. International Journal of Cyber-Security and Digital Forensics (IJCSDF) **7**(1), 87–98 (2018)