19th ICCRTS

"C2 Agility: Lessons Learned from Research and Operations"

Providing Focus via a Social Media Exploitation Strategy

Number: 039

Primary topic: Topic 8: Social Media

Bruce Forrester
Defence R&D Canada – Valcartier
2459 Pie-XI North
Quebec, QC, G3J 1X5
Tel.: (418) 844-4000 #4943
Bruce.Forrester@drdc-rddc.gc.ca

Abstract

For the most part, intelligence agencies have been mere by-standers in the realm of social media (SM). The vast majority of studies and intelligence reports have been post hoc; few if any have been conducted in near-real time. In interviews with intelligence analysts, it was stated that some real-time analysis was conducted simply by using personal twitter accounts for situational awareness. Fraught with security risks and methodological problems, this manner of analysis is not recommended. Contributing to the problem in dealing with SM are the facts that this domain is constantly changing and that there is massive amounts of potential data. Analysts are overwhelmed. Given limited budgets and resources, it is not evident what set of tools and methods can best be brought to bear. This paper presents a possible solution in the form of strategies that will help focus effort and position analysts to build expertise.

Introduction

History has shown that we cannot effectively prepare for the future conflicts by solely applying the lessons learned from the most recent conflict. By today's standards, there were very modest changes that occurred in technology between WWI and WWII. However, the Maginot line, which was definitely built based on the WWI lessons learnt, is a poignant example and proved most ineffective in WWII. As technology, and the affordance that is provided, changes at ever increasing rates, we would also be unwise to concentrate too closely on how social media technologies were used, even a few years ago.

The domain of Social Media is vast and hugely complex; after all we are talking about billions of individuals and organizations contributing to millions of conversations and stories using text, images, video, audio and other media, often all mixed together. The feeling of being overwhelmed is common when one thinks about how to exploit this domain for intelligence purposes. Personally, my thinking has evolved quite extensively in this area. I started off by thinking that we would need teams made up of collectors, analysts, developer/scripter-coders, linguists and cultural specialists, anthropologists and social behaviourists. Such teams would need to be put together for each region of interest that differs from one another in a significant socio-cultural way. Sufficient read-in time and ongoing monitoring would be required by this team. Then the reality of available resources and military budget cuts hit; another way would need to be found. However, this would still be the most comprehensive way if resources were not a problem.

The problem needed to be approached from another angle. What are the unique benefits that SM brings to an intelligence problem? How could adequate coverage of SM sources be accomplished with limited resources? Are there methods and tools that could provide us a head start to analysis when regions heat up? How do we ensure a fast ramp up for analysts learning about new social media platforms and their related analytical tool?

What are some of the solution ideas in this problem space? Significant resource constraints would mandate the need for some form of automated monitoring and alerting using measures tailored for SM and providing indications and warning. Understanding the nature of SM in our regions and countries of interest would allow for a quick decision on the role social media analytics would play as operations were being planned. And then there is the tricky the problem of how to dive deep in specific areas. Are there are common tools, like text analytics for content analysis, which could be easily brought to bear for other types of analysis? It was quickly discovered that there are many tools [1-3] and methods [4-6]. Popular free tools exist but each does a couple of things well, usually on a specific source such as Twitter [7-11] or on an aggregated data set from many social media sources [12-16]. Would this be sufficient and effective enough for the analysts' and commanders' requirements?

The rest of this paper is organised as follows. Section I examines social media as a potential source for intelligence products. Here we look at some previous studies and frameworks that slice and dice the complex social media domain and allow us to think about which aspect might be important. Next, earlier identified [2] methodological problems are re-examined and updated from a practical point of view. Section II presents a subset of results from analyst interviews concentrating on the reasons why social media would be a good source for intelligence. Finally, section III takes the above two sections into consideration and proposes a practically oriented social media monitoring and analysis strategy for resource-limited intelligence communities.

Section I

Monitoring Social Media as an Intelligence Source

Monitoring is a technique used to observe or keep track of important objectives. Intelligence communities use the term surveillance which involves the monitoring of behaviors, activities or things that change over time for the purpose of influencing. With respect to social media, "monitoring is an active monitoring of social media channels for information about a company or organization, usually tracking of various

social media content such as blogs, wikis, news sites, micro-blogs such as Twitter, social networking sites, video/photo sharing websites, forums, message boards, blogs and user-generated content in general as a way to determine the volume and sentiment of online conversation about a brand or topic."[17] Social media presents many challenges when it comes to monitoring. Perhaps the first question to answer is that of focus – what is out there of importance for our needs?

This section will recap some initial research that was conducted looking into the usefulness of social media sites for intelligence. Also examined will be some social media categories and their potential intelligence purpose.

A study [18] of social networking technologies in a counter-insurgency context examined 21 different social networking categories (e.g. blogs, streams, forums, etc.) and how these technologies could be leveraged by insurgents as well by our allies to counter insurgency. The results showed that social networking platforms (Facebook, Twitter, LinkedIn, Pinterest, Google Plus, Tumblr, Instagram, etc.) stand out as a category of significant interest. Five social media technologies (social networking, blogs, forums, video, and live casting) have high potential for use by insurgents in the areas of sympathy and influence, financing, recruiting, and external communications. Another nine have medium potential in additional areas like information gathering, operational coordination and planning. Interestingly, none of the social networking sites examined has better than a low potential for training purposes; a fact that can be exploited by COIN operations. Training facilities will be actual locations that can be targeted in a kinetic sense.

A follow on study [19] expanded upon this first study and put more context around the use of social media by insurgents. An important framework used is the insurgent wheel, Figure 1. This planning cycle that can be applied to asymmetrical warfare. These represent activities that insurgent groups must undertake if they are to conduct those actions necessary to achieve their objectives. Moreover, these phases and activities are not unique, but are typical of those undertaken by any group involved in organizing and executing operations. Intelligence efforts are typically aimed at identifying activities in each of these phases, and their associated key indicators, to develop a series of "signatures" that point to the group's intentions. This in turn may point toward potential insurgent targets and what stage they are at in preparing to act. Nevertheless, it should be noted that while broadly similar in intent, these indicators and signatures will vary substantially from group to group and by region/country.

It turns out that the phase of "Organize" on the insurgent wheel has the most number of probable social media categories available for use and social networking was ranked as high in both activities in this phase. Video is the most diverse category scoring 5 (high potential for use) in the "Organize", "Recruit", and "Act" phases. It would be used for influencing population sympathy, helping to recruit and for promoting the cause after acts have been committed.

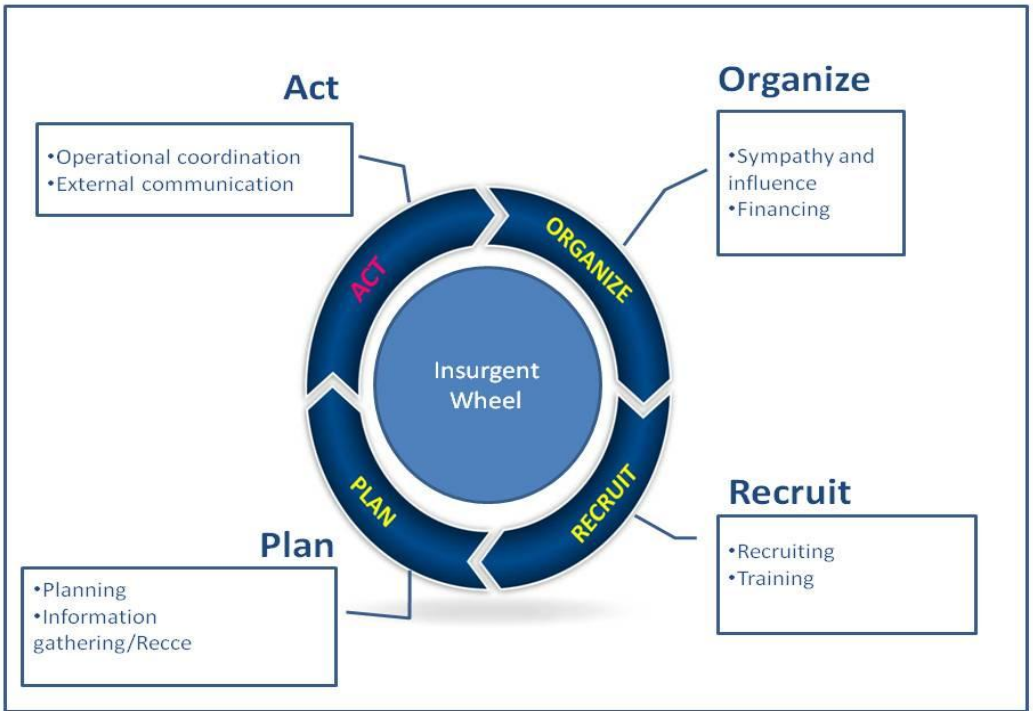

Figure 1: Insurgents activities of interest and the areas of investigations

Perhaps the most significant finding of this study is that the "Organize" phase has the most number of social media categories available for use. Placing emphasis and resources on this phase, before the deployment of troops, makes sense as this is the best place to intervene in a potential insurgency. While the insurgent case was used in this past research, the insurgent wheel is generic and can be used to describe the phase and steps that would be required of any organization conducting an act. So the above results are likely applicable to more general cases, for example a state's military would need to use the same basic steps for force generation and deployment.

The insurgent wheel is only one way to think about social media as a source for intelligence. Often analysts organize sources with respect to their usefulness for a specific requirement. For instance, micro-blogs such as Twitter that are constantly being updated by users around the world provide an excellent near-real time sensor. This sensor can be used to alert analysts to many things such as new events, changing

attitudes, and trending phenomena. It can be used in remote locations and in countries that have been closed off by their governments. Table 1 shows some of the possible intelligence uses of specific social media platforms.

Table 1: Mapping of Intelligence usage to source

Intelligence usage	**Social media source**
• Near-Real time Situational Awareness • Trend watch • Early warning and indicators • Alerting service • Threat assessment	Micro-blog (e.g. Twitter) News feeds
• Targeting (non-kinetic) (i.e. profiling); identifying and getting information about particular person of interest, groups, organizations. • Social Network Analysis	Social networks Genealogy Location Event posting
• In-depth content about a subject or content Profile • Structure of orgs • Understanding the ideology	Blogs Collaborative projects (e.g. Wikipedia) Forums Content communities (e.g. comments on YouTube) Document and Presentation sharing sites
• Collateral damage assessment • Targeting • Standing products • Basic intelligence (baseline) • Country reports	Images (e.g. Instagram, Militaryphotos.net) Video (e.g. YouTube)
• Social Network Analysis • Structure of orgs • Understanding the ideology	Virtual game-worlds (e.g. World of Warcraft) Virtual social worlds (e.g. Second Life)

More will be presented on social media as an intelligence source in section II where analyst interviews are discussed. Another way to see the SM sources is by the quality of content. In looking at the qualities of self-disclosure (what individuals are willing to talk about) and the media richness (relates to required bandwidth and the need for greater exposure in order to understand environment) we can see from Table 2 that blogs and micro blogs offer a high self-disclosure for a "low price" in terms of analyst training time and IT resources. Such qualities make blogs and microblogs an

interesting point of focus. Still, we must be cautious to ensure the truthfulness and validity of the content.

Table 2: Media richness vs Self-disclosure of select social media [20]

	Media richness		
Self-disclosure	Low	Moderate	High
Low	Forums, Wikis	Video and image sharing communities	Virtual games
High	Blogs, Microblogs	Social Networks	Virtual worlds

Methodological Issues

As reported in Twitter as a Source for Actionable Intelligence [2] there are a multitude of methodological issues surrounding social media data, and in particular, twitter (microblog) data. While identified as a good focal point in the above section, we need to ensure that quality of data collected would allow for reliable and valid conclusions to be drawn. Let's examine the previously identified issues and their possible solutions.

The first issue identified was the multiple ways in which researchers collected their data. There was no common data collection method used by the authors of the 40 plus papers examined [2]. In the early days of Twitter, access to all data was available and hence researchers were able to collect and store this data for analysis [21, 22]. Currently, the free API access only provides a sample of tweets [23]. In addition, data collected by way of API is not repeatable as the APIs are not clearly defined and are susceptible to change at any time [24]. This does not necessarily pose a problem if the aim of data usage is for looking at trends across a wide range of subjects. API usage is likely adequate for maintaining situational awareness at a high level. The APIs generally provide all the tweets up to 1500 per minute. If the tweet rate is greater than that it only provides a sampling. So by limiting the search terms to concentrate on areas of interest, it is likely that 1500 would be sufficient even for monitoring. An assessment would be needed for diving deep into the data for detailed analysis. One might require access to the "fire hose" of data which is only available through paid subscriptions. Companies like GNIP [25] Datasift [26] offer complete "fire hose" data (and complete archive of historical Twitter data) from a multitude of North American based social media platforms. Any serious social media analytical capability should

have access to the entire feeds from platforms of interest when needed. In the intelligence case, this "fire hose" access would also be required for platforms that are used within the countries of interest, which in most cases are not available from GNIP.

The second issue [2] dealt with the problem of defining the required target population. While a significant number of the world's population has access to the Internet, only a subset use social media and of that a smaller subset uses twitter [27] and other social media platforms. So knowing who one is looking at becomes very important when reporting for intelligence purposes. Luckily this is of also significant concern to marketer and hence, the business world is working hard on this problem. One way to understand who is using the platform is to mine the profiles provided by the various platforms. At a minimum, users require a username, but there is usually a large set of tombstone data that users can enter voluntarily. For platforms where users are interacting with family and friends, one can expect that the user data would be accurate. These types of platforms could be used as reference points for verifying identities, assuming that usernames are the same or similar. Rao and Yarowsky [28] looked at the ability to detect latent user-properties within social media (age, region of origin and political orientation) with success. As well, Chen et al. [29] examined the "internet water army" or online paid posters in China and found several ways of easily detecting fraudulent usernames and practices. Caution must still be exercised as a study by Cornwell and Lundgen [30] compared misrepresentation in romantic relationships in cyberspace vs. real space and found that people were much more likely to misrepresent themselves in cyberspace. However, understanding the target population requires more than just demographic data.

It is important to understand how the data provider is "packaging" the data. Data providers are business oriented (helping a business find out how their brand is being discussed on social media) and hence, tend to aggregate all of their data sources, further confusing the matter. Good research methods demand the ability to describe the population used for a data sample. This allows precision in building a model or making predictions using a similar population. Nevertheless, depending on the how this data is being interpreted and used, there may be room for compromise. There are hundreds of social media sites and at least 28 different categories of conversation types or purposes [31]. The types of analysis services and tools looked at in [3] used data that had been aggregated from multiple sources without necessarily considering the purpose for which postings (tweets, updates, blog entries etc.) were made. It is highly probable that individual postings were made within a certain context for a certain audience. In their aggregation with other similar topics, which potentially are centred on different time periods or audiences, interpretation and contextual errors easily occur

[32]. If we agree that, in most cases, "where there is smoke there is fire", then an aggregation of platforms and users would not necessarily be a problem for identification of trends. In fact, it might be preferable, allowing for more sensitive thresholds for the indication of change. However, if one is interested in the reactions from refugees and inhabitants of neighbouring countries of an operational area, it is clear that much greater stratification would be needed. In this case geolocation would be most important. As an example, there are large areas of the world that are covered by one dominant social network site, Figure 2, but there is no guarantee that the population of interest will be wholly located in that space. There is research that discusses the problem [33], cases where good stratification of users was possible [27], and ways to detect approximate location through various combinations of pattern of life analysis (active period trends) [34], using contextual clues in user content, comparing time related content to timestamps, use of language, or if a user has several different social media accounts, information from another account might provide location [35].

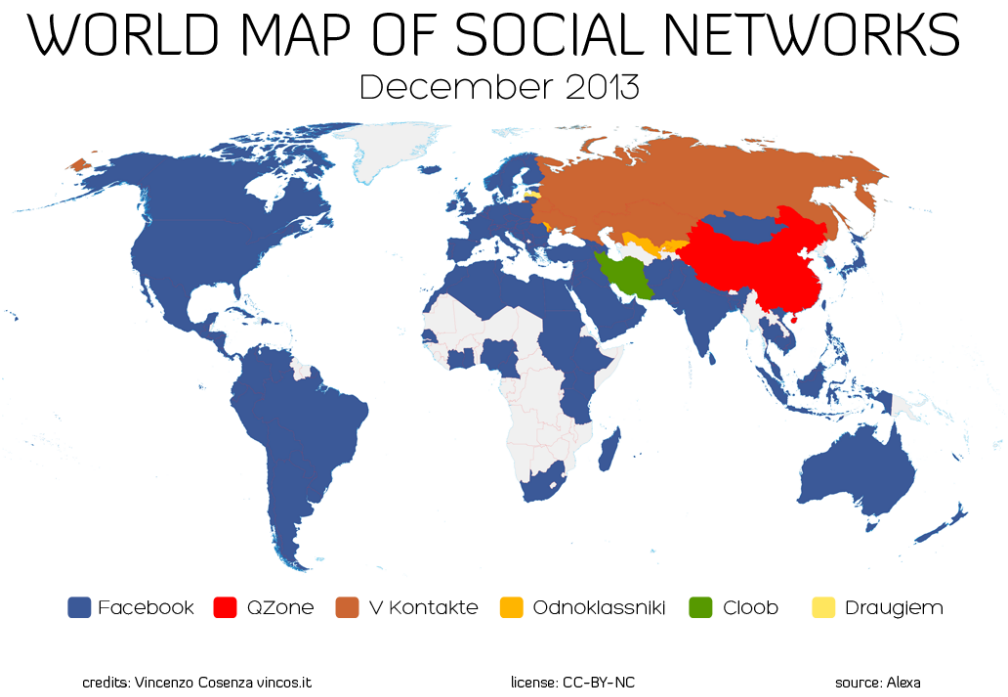

Figure 2: World Map of Social Networks [36]

Social media platforms are in constant change. There is a flow of users joining and leaving for other platforms. There are changes to features. There are buy-outs and amalgamation of platforms. For intelligence work, it is more important to find the social media platforms used by the population of countries of interest Table 3. These tend to

be local instantiations of Facebook or Twitter, or copy-cat platforms. For instance, Weixin [37] and Weibo [38] are two large social media platforms based out of China. "Weixin is relatively similar to Facebook, but it has the option to share to smaller groups of friends and family rather than everyone on your friend list. Weibo is geared more towards the general public and has a more Twitter-like feel" [39].

Table 3: A Sample of Global Social Media sites as of 2011 [18]

Name	Country of origin	URL	Comments
Badoo	UK	http://www.badoo.com/	Over 180 000 000 users (2013 stat.) Although available in most of the world, the site is most active in France, Spain and Italy, as well as in Latin America.
Cyworld (싸이월드)	South Korea	http://cyworld.co.kr/	Around 24 000 000 users in South Korea, China and Vietnam.
Friendster	USA	http://www.friendster.com/	Over 115 000 000 users. The top 10 countries accessing Friendster, according to Alexa, as of May 7, 2009, are the Philippines, Indonesia, Malaysia, Singapore, Pakistan, the United Arab Emirates, Sudan, South Korea, Bangladesh and India.
Hi5	USA	http://hi5.com/	Over 80 000 000 users. Popular in India, Mongolia, Thailand, Romania, Jamaica, Central Africa, Portugal and Latin America.

Name	Country of origin	URL	Comments
iWiW	Hungary	http://iwiw.hu/	On December 24, 2008, the number of registered users on iWiW was 4 million, covering almost all Internet users in Hungary.
Kaixin001 (开心网)	China	http://www.kaixin001.com	Number of users unknown but ranks 135 on Alexa.
Kaskus	Indonesia	http://www.kaskus.us/	As of April 22, 2010, Kaskus has more than 1 620 000 registered accounts. Only available in Indonesian.
LiveJournal	Russia/USA	http://www.livejournal.com/	More than 17 000 000 users in Russia and among the Russian-speaking diaspora abroad. Blogging platform.
Mixi (ミクシィ)	Japan	http://www.mixi.jp/	Invite-only with around 25 000 000 users.
NetLog	Belgium	http://www.netlog.com/	Around 80 000 000 users. Mostly popular in Saudi Arabia, the United Arab Emirates, Yemen, Belgium and Iraq.
Orkut	USA	http://orkut.com/	At least 120 000 000 users (2013 stat.). Owned by Google but very popular in Brazil (53,6% of traffic) and India (35% of traffic)

Name	Country of origin	URL	Comments
P1	China	http://www.p1.cn/	More than 500 000 users in Beijing, Shanghai and Hong Kong. Focus is on upper class elite. Users become members by invitation only.
Qzone (QQ 空间) or QQ	China	http://qzone.qq.com/	At least 1 230 000 000 monthly users (2013 stat.). Important information: A rumour is circulating that Tencent, creator of QQ, is developing a new international SN.
Skyrock	France	http://skyrock.com/	Alexa consistently ranks Skyrock among the top 10 sites in France, Belgium and in Switzerland.
Vkontakte (ВКонтакте)	Russia	http://vk.com/index.php	Over 210 000 000 (2013 stat.) users in Russia, Ukraine, Belarus, and Kazakhstan.

While the methodological issues identified above make it hard to combine studies in this area in order to produce more comprehensive theories, there may be less of an issue for intelligence work provided analysts have a good understanding of the limitations.

Section II

Interview results

As part of an Applied Research Project looking into Social Media Analytics for Intelligence, interviews for analysts' requirements were conducted. This project has as an objective to understand the strategic, operational and tactical intelligence uses of

social media in order to increase the intelligence analysis capability. The goal will be to guide the procurement of appropriate COTS tools or the development of prototypes in the form of automated tools and algorithms for monitoring, capturing, and analysis of data collected from social media for intelligence production.

As part of this research, two sets of interviews sessions were held over a 3-day period in July 2013 and a 2-day period in October 2013. The interviews were designed to capture analysts' opinions and views concerning social media as a source for intelligence. The questions covered importance of social media, the collection, analytical methods and tools as well as some general questions with respect to existing intelligence domains. All interviews took place in unclassified spaces. The interview sessions lasted from 1 to 1.5 hours. A wide range of collectors and analysts were interviewed that covered strategic, operational and tactical oriented personnel. In addition, OSINT, cyber, as well as technical specialists were interviewed. A subset of these interviews is summarized below for selected questions.

Do you think social media as a source is important for Intelligence?

There was a resounding "yes" for this question; absolutely for some. The major reason was as a new source for Situational Awareness (SA) at all levels. It is a space where millions of people are discussing and posting, especially the young, and analysts felt that it could not be ignored. Currently, the way commanders are using INT is currently undergoing a transformation. Analysts are now being asked to identify changes at the strategic level but commanders want this info at a "tactical pace". They felt that Open Source is quick to access and allows one to corroborate the information provided by classified sources. Further, working on the classified side often takes too long due of the many steps required and security issues. Hence by using open sources, one can produce a good report more quickly and "a good report now is often better than excellent report later". However one can't rely on SM solely and we need to ensure SM sources are used along with other sources.

It was stated that social media could be used as a starting point to help guide more intensive and focused collection activities. Social media sources allows analysts to gather other characteristics of communications that were not previously available such as sentiment, mood perception, monitoring for influence and could provide more details for pattern of life analysis.

Here is a compiled analyst's list of possible ways that SM could be used:

1. Monitoring of remote location events (for validation purposes) - YouTube has proven to be very useful and used to complete INT reports about damages and armaments used.
2. Follow tweets from influencers.
3. Perceive sentiment about an evolving conflict situation (predict what's next)
4. Perceive reaction and local population attitudes about some decisions on foreign policies.
5. Quickly analyse the effect of operations on local population (assess kinetic effect, measure success, understand social reaction to operation, define influence requirements).
6. Corroborate quickly some information from other INTs (need many sources).
7. Identify the real intent of individuals and groups (identify contradictory discourses) - variations of intent.
8. Identify patterns of life of individuals.
9. Identify, understand, and monitor social networks.
10. Dig information triggered by SIGINT activity.
11. Perceive the extent of misinformation about a situation.
12. Perceive the difference between what media is advertising and what people are thinking (on what side do they stand).
13. Stay up-to-date about how people exchange information.
14. Monitor influence - Identify both strong leaders and soft leaders in the SM world.
15. Force protection (monitor exposure of critical military assets and operations to outside world).
16. Rumor detection.
17. Monitoring "friendly" SM for Counter Intelligence, OPSEC and the adversary's perception of us.

What would be the differences between Strategic, Operational and Tactical use of this source?

In general, SM sources can be useful at all levels where SA is required. It could provide input and indicate intent from a strategic point of view. As such it could be used to identify what is trending (i.e. on Twitter). Keeping secrets (i.e. by an oppressive government) is harder than ever before because of the pervasive nature of social media, although there still are examples were tight internet control exist. By using social media in a strategic role, one could determine how many times a search term or keyword was used. This could then be used to determine the impact on the general public of some decisions made by governments. An example was given for strategic use where a certain country published White Paper in a prominent American newspaper and monitored social media for the reaction.

For tactical, SM could provide an "insider's view" of some locations that are hard to reach without being there or which are closed to outsiders. An identified tactical use

was to confirm and validate collateral damage following air strikes in Libya. Locals would use their mobile devices to film bombed areas and post directly online to sites like YouTube. The coalition task force would then have another source to verify against. This type of reporting as well as textual accounts of events inside countries, in effect, provides a double check capability. Further one can compare official government reports to those on SM.

What type of questions can be asked for which social media might help answer? What is it that you are trying to understand and/or uncover?

The main interest from analysts for social media was to help answer questions that require situational awareness; what's happening in a certain location right now? However, most analysts stated that they currently do not receive many SM related questions – due in part to the novelty of this source. Further, collectors and analysts have limited access to SM and the necessary tools to do the monitoring.

Analysts felt that SM could be used to measure the effect of actions. Again, a use would be to look for videos to confirm classified sources and see if what is reported is real (as in collateral damage reports from Libya where this was very useful).This would not be limited to just the kinetic information about the effect, but the social effect and perception from both local and worldwide could be monitored via SM. There was reported a need for a capability to understand how people, both local and internationally, perceive an action. SM could provide the data and used to measure the response to events at different levels, and by groups.

It was stated that the analysts (and commanders) require an understanding of what social media can help answer form an intelligence perspective. The analysts sensed that SM sources would be useful for some situations where no other information sources are available (e.g. Syria), or when there is interest in a meeting somewhere and you would like to know who was attending. Finally, the conjectured that SM could help to understand intent.

What stands in the way of successful use of social media as a source?

By far the largest obstacles mentioned were Information Management and Information Technology policy and security issues. For security reasons, there is limited access to the open Internet on military networks. Analysts require specialized tools, which allow for language translation, and further training in the domain. Currently some analysts and collectors have used personal accounts to access SM sites. However they stated that there is a real danger of being tracked back to a personal IP address. Further,

there is a great danger if analysts use their personal accounts for collection. Hence, work locations are used to access SM data due to these reasons.

There is also a wide held belief amongst commanders of all nations that reports based on unclassified intelligence is not as valuable or reliable as classified information. Thus there is not the same demand (despite that analysts claim that in many INT products they used upwards of 85% of intelligence that was gathered from open sources). One analyst stated that it is "important not to burn your credibility in the branch so we take care with social media in our reports."

It is very hard to find the time needed to initially learn about and then stay up to date on tools, methods, and the ways people exchange info on social media (online gaming, chat …). In addition, analysts are not allowed to use these sites at work for personal reasons; which would provide a greater base of understanding of SM. Analysts stated that it is difficult for them to actually follow someone; rather they must ask for help from the collectors.

As social media is a relatively new source, the INT community has not yet determined how to use, validate and corroborate with the other INT sources. There is a lack of knowledge and tools for the filtering of huge amounts of social media data. How much data do we need to keep? Do we need or want to look back in time or just look at recent streams of data? Is the way terrorists use Twitter different than tweets from ordinary people? There is a significant lack of knowledge on how to exploit social media.

Discussion

Thousands of social media sites, billions of posts, and the lack of access, tools and methods have all lead to analyst feeling currently overwhelmed. As a result, there has been very little movement towards a comprehensive approach for the monitoring, analysis and exploitation of social media as a source for intelligence. This despite analysts' collective view that social media is a crucial area and one which is increasing in importance.

In summing up the interviews, it can be said that analysts believe that social media can be a valuable source of information and data for intelligence reports. They are currently looking to social media data for situational awareness in main part due to the fast pace of reporting that is now common to intelligence. They felt that there is large potential for other types of analysis that fall into one of two categories: 1) trend and sentiment analysis for situational awareness; and 2) deep dive analysis, that looks at specific entities in great detail.

Analysts believe that SM data can be useful from tactical to strategic purposes. However there are significant challenges in setting up a SM program, for example: 1) Information management and information technology policy and security issues need to be challenged and adjusted to ensure wider access to SM; 2) Anonymous access (as much as is possible) must be brought close to analysts' workspaces; 3) Attitudes in the Intelligence Community will need to be adjusted to see the value of SM assessments; and 4) Tools, methodologies, and training will need to be acquired or developed.

When planning for the exploitation of social media, there are some very practical aspects that must be considered. With respect to capacity, military budgets are being cut back around the world leaving a relatively small number of analysts and few resources internally for outsourcing. With respect to SM platforms, there are many different platforms used around the world and these are constantly changing in features and users. With respect to data, there is a huge amount that has already been produced and the growth of data is astounding. These combined leaves analysts feeling dazed. There must be focus and clarity in order to move forward in an effective way.

Section III

Description of a Strategy for SM Monitoring for Trend and Opinion Analysis

Let's look at a strategy for the monitoring and analysis of SM for situational awareness. It covers the first of the two types of analysis desired by analysts: trend and opinion (sentiment) analysis.

The interviews established that analysts need an ability to produce fast SA for commanders. As we saw from Table 2, blogs and microblogs offer a high self-disclosure for a "low price" in terms of analyst training time and IT resources. Table 1 also showed that blog and microblogs are best suited to near real-time situational awareness. In fact, microblogs are probably best as they are fast moving and reflect the opinions of people right now. A baseline technique would need to be established and thresholds developed for trend analysis. Blog and microblogs are both easily available from APIs and websites. There are also many tools that monitor for trends that are relatively easy to learn. However, unless developer support is readily available, it is best to use a service that will ensure the maintenance of needed APIs and incorporation of new metadata and features as SM platform change. With some development, automation can easily be applied to monitor trends of interest and provide alerts when pre-established thresholds are reached. How would this look?

In earlier research [1] form 2012, it was reported that the immediate advantages of established COTS based monitoring services are that one can easily find a service that provides comprehensive monitoring coverage of SM. Services provided an access to multiple sites. The services were not very expensive and were simple to set up and use. However, there were also some significant disadvantages identified at that time. These companies were built for commercial use and had little to no analytical tools built into their services, data cleaning was hit and miss, there was very limited access to foreign platforms, and finally the big security concern of who could see the searches and keywords used.

Two years later and many of these concerns have been "overrun by events". There are at least two comprehensive data providers [25, 26] that have a large selection of SM sources and continue to acquire more sites. These data providers have the entire data set from Twitter and hence can be used for data and long-term trend analysis. They provide some ability to clean data. These data providers can be used to feed other services [40-42] that offer greater ability to do business intelligence, monitoring, sentiment and trend analysis. Further, if you use these services, there are ways to easily export the collected data to be further analysed by custom tools at your own location. Security remains an issue, but the data is internally stored by these companies and hence searches are not made public. Agreements with specific companies can increase security issues further. Of course if one could afford to, one could simply buy all of the data direct from the data providers and import that into a secure internal server to be much more secure; Good luck with that ☺ – an external service offers the biggest bang for the buck and can be set up to be relatively secure.

So, trend monitoring and basic analytics (sentiment, semantic, demographics, and location) can be effectively handled by paid services. Analysts would set up searches using blog and microblogs feeds as data based on their area of interest. COTS services generally do not allow the automatic ability to set thresholds for warning and automatic alerting. For now, regular exports of data will need to be made and fed into services that can perform these functions.

Description of a Strategy for SM Deep Dive Analysis

The second area where analysts felt that there was large potential was deep dive analysis. This is a much harder problem given the constraints already mentioned above. Specifically, the greatest problem is the constantly change to 1) the ways the SM sites are being used, 2) the SM platforms features and interfaces, and 3) the content itself (much shorter lifetime of relevance). This last one is not specific to SM

but will, for example, affect the ability of automated tools to find entities. We need to ask, are there any relative constants in all this change?

It turns out that there is a significant amount of metadata that can be collected about a countries use of social media [36, 43-48]. Such metadata, see Table 4 for an example of potential areas, can then be used to determine firstly, the need for SM analytics with regard to a country of interest, and second, the immediate requirements for analyst expertise. The deep dive analysis strategy involves collection and analysis of critical metadata about the SM usage similar to an intelligence country report for other aspect such as geography, politics, economics, etc. When a region heats up, analysts will be able to refer to the applicable SM country profiles to determine the importance that SM will have in the intelligence collection, planning of the operation, and for tactical purposes. Such metadata is likely to change at a much slower rate than the actual content produced on the SM platforms and will need only periodic updates. In some cases, the rapid change in certain metadata could also indicate something interesting such as when thousands of people joined social networking site during the Arab Spring [27].

Table 4 Example of the metadata collected on a specific country of region

Country of interest:
1) What are the main social media platforms being used? a) Types most frequently used (blogs, video, image etc.) b) Who are using each? i) Age ii) Sex iii) Religion c) Method of access
2) What are the main topics of interest discussed in the country a) Blogs b) Mico-blogs c) Videos
3) What countries interact together in discussions? a) What countries do followers come from? b) What other countries are most followed by people?
4) Who has control of these platforms? a) What is the level of government monitoring of these platforms? b) What types of actions are taken against SM users?
5) Geolocation data a) What other countries are sharing these platforms? b) What are the main other countries that contribute for the popular discussion for the country of interest?

6) Cyber concerns
   a) What types of deception originate from this country?
   b) Use of bots
   c) Criminal activities

Another key aspect to this deep dive strategy will be identifying the personnel and training required such that deep dive analysis can be conducted. It is unlikely that analysts will be required to do such analysis for the entire set of possibilities at the same time. There will also be a large overlap in the types of analytical tool that can be used to make sense out of the data. Hence, there will be a basic course that can be offered that will allow for a quick ramp up of skills required for the delta between an analyst current ability to analyse SM sources and the specific SM sources used in the country of interest. Monitoring for and preparing to teach the delta could be assigned to one or two people. They might also be responsible for the SM country profile reports thus maximizing the use of resources.

So in summary, Social Media Country Profiles, Table 4, could be generated for all countries of interest. The maintenance of such profiles will allow for fast ramp up when a country of interest becomes a potential operational area. The profiles will allow for analysts to quickly assess the usefulness of SM as a source depending on the nature of the operation. If SM is deemed useful, analysts can quickly be trained on the idiosyncrasies of the SM platforms used in that country and region, allowing for granular analysis.

There are still many aspect of analysis that will need development and integration with the monitoring techniques. At a basic level, specific SM platforms will need to be connected to using APIs and the nature of the data and metadata analysed. The interpretation of content in multiple languages is getting easier, but the technology is still not fully mature and the cultural meaning are farther behind. The types and combinations of deep dive analytics will need to be determined and adapted to SM data.

Conclusion

This paper looked at issues and challenges surrounding social media as a potential source for intelligence products. It started with a review of studies and frameworks that allowed for simplifying the complex domain of SM. A practical point was taken to re-examine and update methodological problems. Next analyst interviews were presented and provided evidence as to why social media would be a good source for intelligence. The combination of the study insights and the interview results lead to a

practically oriented social media monitoring and analysis strategy for resource-limited intelligence communities.

Granted there remain interesting challenges with the integration and development of tools. At the fore front will be the combining of automated tools, that provide alerts of significant change, and the metadata reports, that will provide sufficient insight into how and if social media can be useful for a given situation and its related questions. As well, there is a need to identify and implement the deep dive analytic tools and methods. Nevertheless, it is felt that a focus has now been provided. Action can be taken immediately and plans can be made to surmount the remaining challenges that will lead to the accomplishment of the goal: the ability to monitor SM for pertinent changes, based on commanders' intelligence requirements and to be able to provide deep dive analytics for intelligence production.

References

[1] Forrester, B., *Social Media Exploitation Tools: Understanding Where and How to Look*, in *HFM-201 Specialist Meeting on Social Media: Risks and Opportunities in Military Applications*, N. RTO, Editor 2012, RTO NATO: Tallinn, Estonia.

[2] Forrester, B. *Twitter as a Source for Actionable Intelligence*. in *18th Command and Control Research and Technology Symposium*. 2013. Alexandria, Virginia: Command and Control Research Program.

[3] Labrèque, A., *Study of Social Networking Exploitation Tools*, B. Forrester, Editor 2011, Defence Research and Development Canada: Quebec City.

[4] Adedoyin-Olowe, M., M.M. Gaber, and F. Stahl, *A Survey of Data Mining Techniques for Social Media Analysis*, in *arXiv.org*2013: Cornell University Libary.

[5] Fan, W. and M.D. Gordon, *Unveiling the Power of Social Media Analytics.* Communications of the ACM, In Press(June 2014): p. 26.

[6] Pang, B. and L. Lee, *Opinion mining and sentiment analysis.* Foundations and Trends in Information Retrieval, 2008. **2**(No 1-2): p. 1-135.

[7] *Tweriod* 2014; Available from: http://www.tweriod.com/.

[8] *Followerwonk*. 2014 [cited 2014 4 April]; Available from: http://followerwonk.com/.

[9] *Twitonomy* 2014 [cited 2014 4 April]; Available from: http://www.twitonomy.com/.

[10] *Twilert*. 2014 [cited 2014 4 April]; Available from: http://www.twilert.com/.

[11] *TweetDeck*. 2014 [cited 2014 4 April]; Available from: https://about.twitter.com/products/tweetdeck.

[12] *Klout*. 2014 [cited 2014 4 April]; Available from: http://klout.com/home.

[13] *Hootsuite*. 2014 [cited 2014 4 April]; Available from: http://try.hootsuite.com/.

[14] *SocialFlow*. 2014 2104]; 4 April]. Available from: http://www.socialflow.com/.

[15] *SocialBro*. 2104 [cited 2014 4 April]; Available from: http://www.socialbro.com/.

[16] *ArgyleSocial*. 2014 [cited 2104 4 April]; Available from: http://argylesocial.com/.

[17] Wikipedia. *Social media measurement*. 2014 [cited 2014 17 March]; Available from: http://en.wikipedia.org/wiki/Social_media_monitoring.

[18] Labrèque, A., *Study of social networking technologies Social networking analysis in a counter-insurgency context*, 2011, Defence R&D Canada – Valcartier: Quebec City.

[19] Forrester, B., A. Frini, and R. Lecocq. *Understanding the Role of Social Media in a Counter-Insurgency Context*. in *NATO IST-099 RSY-024 Emerged/Emerging "Disruptive" Technologies*. 2011. Madrid, Spain.
[20] User, A., *GEOINT: applications, challenges, and capabilities*, in *Big Data Conference*, TTC, Editor 2012: Arlington, VA.
[21] Cha, M., H. Haddadi, F. Benevenuto, and K. Gummadi. *Measuring User Influence in Twitter: The Million Follower Fallacy*. in *ICWSM '10: Proceedings of international AAAI Conference on Weblogs and Social*. 2010.
[22] Byun, C., Y. Kim, H. Lee, and K.K. Kim. *Automated Twitter Data Collecting Tool and Case Study with Rule-Based Analysis*. in *iiWAS2012*. 2012. Bali, Indonesia.
[23] Twitter. *REST API Rate Limiting in v1.1*. 2014 [cited 2014 2 April]; Available from: https://dev.twitter.com/docs/rate-limiting/1.1.
[24] Black, A., C. Mascaro, M. Gallagher, and S. Goggins, *Twitter Zombie: Architecture for Capturing, Socially Transforming and Analyzing the Twittersphere*, in *GROUP`12*2012: Sanibel Island, Florida.
[25] GNIP. *GNIP Social Media data provider*. 2014 [cited 2014 24 March]; Available from: http://gnip.com/.
[26] Datasift. *Datasift website*. 2014; Available from: http://datasift.com/.
[27] Howard, P.N., A. Duffy, D. Freelon, M. Hussain, W. Mari, and M. Mazaid, *Opening Closed Regimes What Was the Role of Social Media During the Arab Spring?*, N.S. Foundation, Editor 2011, The Project on Information Technology and Political Islam: Washington.
[28] Rao, D. and D. Yarowsky, *Detecting Latent User Properties in Social Media*, 2009.
[29] Chen, C., K. Wu, V. Srinivasan, and X. Zhang *Battling the Internet Water Army: Detection of Hidden Paid Posters*. eprint arXiv:1111.4297, 2011.
[30] Cornwell, B. and D.C. Lundgren, *Love on the Internet: involvement and misrepresentation in romantic relationships in cyberspace vs. realspace.* Computers in Human Behavior, 2001. **17**: p. 197-211.
[31] Solis, B. and JESS3, *The conversation prism representation*, 2010.
[32] Spark, D., *Real-Time Search and Discovery of the Social Web*, S.M. Solutions, Editor 2009.
[33] Haewoon, K., L. Changhyun, P. Hosung, and M. Sue, *What is Twitter, a social network or a news media?*, in *Proceedings of the 19th international conference on World wide web %@ 978-1-60558-799-8*2010, ACM: Raleigh, North Carolina, USA. p. 591-600.
[34] Eagle, N. and A. Pentland *Reality mining: sensing complex social systems*. Pers Ubiquit Comput, 2006. 255-268.
[35] Cheong, M. and V. Lee, *Integrating Web-based Intelligence Retrieval and Decision-making from the Twitter Trends Knowledge Base*, in *SWSM'09*2009: Hong Kong.
[36] Cosenza, V. *Vincos Blog* 2014 [cited 2014 25 March]; Available from: http://vincos.it/social-media-statistics/.
[37] Weixin. *Weixin Web site*. 2014 [cited 2014 2 April]; Available from: weixin.qq.com.
[38] Weibo. *Weibo website*. 2014 [cited 2014 2 April]; Available from: www.weibo.com.
[39] Totka, M. *Popular Social Networks in Other Countries*. 2013 [cited 2014 24 March]; Available from: http://www.business2community.com/social-media/popular-social-networks-countries-0611780.
[40] *Nexalogy Environics*. 2014 [cited 2014 2 April]; Available from: http://nexalogy.com/.
[41] *Radian6*. 2014 [cited 2104 2 April]; Available from: http://socialcloud.radian6.com/docs.
[42] *Collective Inelligence*. 2014 [cited 2014 2 April]; Available from: http://www.collectiveintelligence.com/Pages/default.aspx.

[43] Solis, B. *The State of Social Media Around the World 2010*. 2010 [cited 2014 2 April]; website]. Available from: http://www.briansolis.com/2010/02/the-internationalization-of-social-media/.
[44] Dennis, E.E., J.D. Martin, and R. Wood *How People in the Middle East Actually Use Social Media*. The Atlantic, 2013.
[45] Consunji, B. *What's the State of Social Media in Asia?* 2012 [cited 2014 2 April]; Available from: http://mashable.com/2012/06/18/social-media-asia-google-hangout/.
[46] Pereira, A., G. Kumamura, C. Pownall, and Z. Nooruddin *Asia-Pacific Social Media infographics H1 2011*. 2011.
[47] PewResearch, *Demographics of Social Media Users*, P. Research, Editor 2013.
[48] Quantcast. *Quantcast*. 2014 [cited 2014 2 April]; website]. Available from: https://www.quantcast.com/.