13.

Countering Lone Actor Terrorism—Weak Signals and Online Activities

by Lisa Kaati and Fredrik Johansson

As has been made clear in the previous chapters of this book, lone actor terrorists can come in a variety of shapes and with a range of backgrounds and they are generally very hard to detect before they attack. As argued by Cohen[1] and in this volume, there are no clear profiles for lone actor terrorists since they have a large variation in factors such as social status, ideology, and personality type. Even though intelligence agencies do their best to keep track of individuals who are violence-prone and have radical ideas and thoughts, it has repeatedly been shown that attacks often come from directions that are not expected beforehand, for example the attacks carried out by Anders Behring Breivik in Norway or the various attacks from born-and-raised Americans in the United States.

In the search for potential lone actor terrorists, it is important to have an open mind. Wilhelm Agrell[2] refers to the black swan theory, a metaphor introduced by Nassim Nicholas Taleb,[3] for describing events that have major effects and come as a surprise to the observer. The black swan theory is according to Agrell an explanation to why the Norwegian security service could not stop Anders Behring Breivik before his attack, and the same argument can be used for explaining the successful execution of many other terrorist attacks. Humans are generally quite bad at coping with black swans and other kinds of low-probability events since we have many cognitive biases leading to the fact that we try to confirm what we already suspect, rather than looking for

[1] Katie Cohen, *Who Will Be a Lone Wolf Terrorist?* (Stockholm: FOI Technical report: FOI-R--3531--SE, 2012).

[2] Wilhelm Agrell, *Den svarta svanen och dess motståndare: Förvarningsaspekter på attentaten i Oslo och på Utøya 22 juli 2011* (Stockholm: National Defence College, 2013).

[3] Nassim Nicholas Taleb, *The Black Swan: The Impact of the Highly Improbable* (London: Penguin, 2nd edn 2007).

alternative hypotheses. Such cognitive biases hold true also for intelligence analysts, making it important to provide them with support for evidence-based reasoning and methodologies that encourage the search for alternative hypotheses.

In this chapter, we will discuss how computerized decision support systems and various computational techniques can be used to analyze and understand weak signals of an upcoming terrorist attack from a lone actor.[4] More specifically, we will mainly focus on how various kinds of social media monitoring and analysis techniques can help in this process. Before discussing this in further detail, we would like to point out that technological solutions alone are no golden bullets that will stop terrorist attacks from happening. What technical solutions potentially can accomplish or add is that they can support intelligence analysts in their work of protecting society by detecting potential lone actor terrorists and thus attempting to stop terrorist attacks before they take place.

Technology nearly always comes with a price and technological solutions for making society more secure are a double-edged sword that may have implications on peoples' privacy. More or less the same techniques that can be used in attempts to protect society against terrorism in a democratic country may be used to monitor the opposition in a more repressive regime. It is therefore of fundamental importance that legal and ethical considerations are taken into account before using technological means for fighting lone actor terrorism. The techniques we will discuss or suggest here are intended to have as low impact on the privacy of individuals as possible, but all kinds of social media monitoring analysis are affecting peoples' privacy, no matter whether the purpose is to discover potential lone actors or to market a new product.

Internet and Lone Actor Terrorists

Internet and social media have an enormous effect on modern society. People can search for and find information fast on nearly any topic, and it is easy to communicate and keep in touch with relatives and friends living far away. It is also easy to discuss and communicate political views and opinions using various forms of

[4] The technical discussions in this chapter are held on a comprehensive and non-detailed level. For more technical details, we refer the interested reader to our previous work: Joel Brynielsson, Andreas Horndahl, Fredrik Johansson, Lisa Kaati, Christian Mårtenson, Pontus Svenson, "Harvesting and Analysis of Weak Signals For Detecting Lone Wolf Terrorists," *Security Informatics* 2: 11 (2013).

social media. Today there exist many different kinds of social media services and new services are constantly being developed.

Several terrorist organizations have a presence on the Internet and are spreading their propaganda through micro-blogs such as Twitter, social networking sites such as Facebook, and various discussion forums. According to many reports, several of the recent examples of homegrown terrorists have been influenced and encouraged by recruiters and motivators such as Anwar al-Awlaki (including Nidal Malik Hasan who carried out the Fort Hood shootings in the United States). The online magazine *Inspire*, produced by Al-Qaida in the Arabian Peninsula (AQAP) is disseminated worldwide using various social networks, blogs, jihadist forums, and file-sharing web sites.[5] *Inspire* is to a large degree focusing on the recruitment and training of young Western Muslims to fight unbelievers by carrying out lone actor attacks. Likewise, videos encouraging terrorism are spread via various social media services such as YouTube and discussion forums. The development of Internet and social media shrinks the world so that it becomes easier to communicate with the like-minded, no matter if the topic for discussion is modern art or how to make a bomb in your mothers' kitchen using household items, as described by *Inspire*.

There are many examples of how lone actor terrorists have been using various forms of social media for communication and inspiration. This is of course unsurprising since very many people make use of social media. What is of interest is that the use of social media and the Internet in many cases leave digital traces that can be gathered and analyzed. The use of encryption techniques and password-secured sites may be used for exchanging sensitive information, but in most circumstances, the actual aim for the individuals or terror organizations is to reach a wider audience, making it likely that those who want to encourage violent extremism will remain using social media services that are accessible by everyone, including the intelligence services.

One example of a lone actor terrorist who used social media is Anders Behring Breivik. Breivik made use of several different social networking sites, including Facebook and Twitter. He also posted the manifesto *2083: A European Declaration of Independence* on the Internet before he carried out his terror attacks in Norway. An analysis of the content posted by Breivik reveals that many of his postings indicated that he had radical beliefs. It should be noted that radical belief is not a crime, but in combination with other activities (such as the acquisition

[5] Edan Landau, *And Inspire the Believers* (Herzliya, Israel: International Institute for Counter-Terrorism, IDC, 2012).

of chemicals that can be used for bomb making) the expression of radical beliefs could have worked as warning signs. Another example is Mohamed Merah who in 2012 killed several Jewish schoolchildren and a Rabbi. Merah used a camera strapped to his chest to record the killing of all his victims and posted the footages online before he was shot to death by a police sniper.[6] Although the postings of Breivik's manifesto and Merah's footages were posted too late to be useful for predicting or revealing the attacks, there are many examples of how people who have carried out lone actor attacks or school shootings have posted revealing content online way before their attacks.[7] If it was possible to detect these postings before the actual attacks the material could have served as weak signals of an upcoming attack. If purely manual means are used, it is highly unlikely to find such content but if semi-automated or automated methods for searching for this kind of material are used, the possibility to detect such materials before an actual attack increases.

Keeping track of what individuals are sending terrorism propaganda and who their followers are is at least in theory a valuable tool for finding out potential lone actors, since even though lone actors are carrying out their attacks in isolation, this does not mean they are not communicating with or influencing each other. In fact, according to Sageman,[8] most lone actors are part of online forums, making digital traces of uttermost importance when trying to identify threats to the society. As stated by Weimann,[9] "In nature, wolves do not hunt alone: they hunt in packs. So, too, with the lone-wolf terrorists: there is a virtual pack, a social network, behind them. They may operate alone, but they are recruited, radicalized, taught, trained and directed by others". Hence, even though they carry out their attacks on their own, this does not mean that they are not communicating with others. In fact, it is often claimed that nearly all radicalization of lone actor terrorists is taking place on the Internet.

Social Media Monitoring and Analysis

Monitoring and analyzing various social media sites has become an important task for many different reasons. By using a variety of state-of-the-art techniques online content from social media services can be gathered and

[6] Gabriel Weimann, "Lone Wolves in Cyberspace," *Journal of Terrorism Research* 3: 2 (2012).

[7] Alexander Semenov, Jari Veijalainen, and Jorma Kyppö, "Analysing the Presence of School-shooting Related Communities at Social Media Sites," *International Journal of Multimedia Intelligence and Security* 1: 3 (2010).

[8] Marc Sageman, *Understanding Terror Networks* (Philadelphia: University of Pennsylvania Press, 2004).

[9] Weimann, "Lone Wolves in Cyberspace."

analyzed. The goal of the analysis can for example be to get information about public opinion on a certain topic or to get information about the members and the sub-groups of a social network. When using social media monitoring and analysis to detect threats towards society posed by individuals, the goal is to identify weak signals indicating that someone is planning a terrorist attack. A weak signal can be seen as an early warning for an upcoming event. There are few events that do not have any prior warning signs; the problem is to find the right signals and to analyze them properly.

The ability to identify weak signals that are present on the Internet requires tools for the monitoring of extremist web sites and social media accounts belonging to or associated with known terrorist groups. Similar tools exist and are currently being used by companies for marketing purposes. We will here not go into any technical details on how such tools can be implemented, but rather briefly describe various functionalities that can be of importance when searching for weak signals that can be used to detect potential lone actor terrorists in social media.

We will only discuss methods that can be used on publicly available web data or social media data that is accessible by anyone interested in the data. Recent research suggests that this kind of publicly available data can be used for many purposes, including predicting the winner of elections,[10] estimating private attributes such as ethnicity or political views,[11] or predicting the stock market.[12] Although claims that the mood people express in tweets can be used to predict the stock market should be taken with a pinch of salt, it is quite uncontroversial that it is possible to find out useful and actionable information from social media.

The main challenges when it comes to social media monitoring and analysis are (1) to collect data or information that may be of potential interest and (2) make further analysis of the collected information. While human analysts are much better at analyzing the actual content of text than machines are, the amount of user-generated content on the Internet grows so quickly that it is impossible for humans to read and process all data.

[10] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe, “Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment,” *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (2010), 178-185.

[11] Michal Kosinski, David Stillwell, and Thore Graepel, “Private Traits and Attributes Are Predictable From Digital Records of Human Behavior,” *PNAS* 110: 15 (2013), 5802-5805.

[12] Johan Bollen and Huina Mao, “Twitter Mood As a Stock Market Predictor,” *Computer* 44: 10 (2011), 91-4.

Various degrees of automatic processing is therefore necessary, although the final assessments and judgments always should be performed by an analyst.

When analyzing social media, there are at least two aspects that can be highly relevant. First, the actual content of the posts is of fundamental importance. Various keyword searches can be used for finding social media posts that contain the specific terms, but there are more sophisticated techniques available for analyzing the content. With topic extraction algorithms, it becomes possible to search for posts relating to specific topics (for instance, terrorism), without having to manually specify what keywords to look for. Furthermore, by using affect analysis algorithms,[13] it is possible to identify posts that contain strongly expressed negative emotions such as hatred or anger.[14] Second, in addition to looking at the content, it can be of interest to analyze the structure of the social network. If we have knowledge about a few potential terrorists or influencers, their friends, followers, and so on, can be identified, in order to analyze who is communicating with whom. Moreover, various social network analysis techniques can be applied to find the most central individuals in the network, making it possible to focus the available resources on those individuals.

Weak Signals

There are many different kinds of weak signals that might precede a terror attack that are possible to detect. For example it might be feasible to detect weak signals indicating that someone has the *intent* to carry out an attack, that someone has the *capability* to carry out an attack, or that someone has the *opportunity* to carry out an attack. Typically, it is of interest to identify weak signals such as someone with radical beliefs and extreme hate, knowledge about how to produce homemade explosives, and interest in firearms and signs of rehearsal (with explosives or shooting). Apart from these kinds of quite concrete signals, it is also possible to search for more complex signals that represent certain warning behaviors.

[13] Fredrik Johansson, Joel Brynielsson, and Maribel Narganes Quijano, "Estimating Citizen Alertness in Crises Using Social Media Monitoring and Analysis," *Proceedings of EISIC 2012* (2012), 189-96.

[14] Ahmed Abbasi, Hsinchun Chen, Sven Thoms, and Tianjun Fu, "Affect Analysis of Web Forums and Blogs Using Correlation Ensembles," *IEEE Transactions on Knowledge and Data Engineering* 20: 9 (2008).

Available literature on lone actors and school shooters shows that there are various warning behaviors that have been empirically proven to precede terrorist attacks and school shootings.[15] Meloy and others propose the following list of warning behaviors that precede acts of targeted violence, relate to targeted violence, or may predict it:[16]

- Pathway warning behavior
- Fixation warning behavior
- Identification warning behavior
- Novel aggression warning behavior
- Energy burst warning behavior
- Leakage warning behavior
- Directly communicated warning behavior
- Last resort warning behavior

Pathway warning behaviors include the planning, preparation, and implementation of an attack. Part of the planning and preparation can be the acquisition of the required knowledge and material to carry out the attack, for instance by searching for and downloading material in order to learn how to build a pipe bomb or ordering fertilizers with which to make explosives. Fixation behavior is by Meloy and others described as behaviors indicating an increasingly pathological preoccupation with a person or a cause, which, for instance, can be recognized as an increasingly negative characterization of the object of fixation, coupled with an angry emotional undertone. Identification warning behavior can for instance be expressed as someone having a warrior mentality, associating closely with weapons, or identifying oneself with previous attackers or assassins. Novel aggression warning behavior relates to behavior that shows the capacity of violence, while energy burst warning behavior relates to an increase in the frequency or variety of activities related to the target when the day of attack

[15] Cohen, *Who Will Be a Lone Wolf Terrorist?*

[16] J. Reid Meloy, Jens Hoffmann, Angela Guldimann, and David James, “The Role of Warning Behaviors in Threat Assessment: An Exploration and Suggested Typology,” *Behavioral Sciences and the Law* 30: 3 (2012), 256-79.

is approaching. Leakage warning behavior is expressed when the intent of carrying out an attack is communicated to a third party, which is similar to the directly communicated warning behavior, except that in the second case this is a direct threat. Finally, last resort warning behavior can be seen as an expression of an increasing desperation or distress, in which the individual sees no way out except by taking violent action.

While some of these behaviors are most likely to be detected using physical interaction, many of the behaviors can also be identified in social media posts or other kinds of Internet-related activities. Some of the behaviors relate to the *capability* to carry out an attack, some of them relate to the *intent* to carry out an attack, and some relate to the *opportunity* to carry out an attack. It is most likely that some of the warning behaviors can be identified by using various automated analysis techniques. More specifically, natural language processing can be used to find out concepts to which the author seems to be fixated, or on which the author expresses a very negative or positive sentiment. In the same manner, by looking for word patterns containing auxiliary verbs signaling intent together with words expressing violent actions, leakage or directly communicated warning behavior may be detected. While text analysis may be very useful, it is also common that images or videos contain clues to whether someone shows warning behaviors. As an example, it is not uncommon that attackers post images or videos where they pose with weapons long before they carry out their attacks. For detecting such information, various content-based image retrieval algorithms can be applied, even though they so far have a too low accuracy to be used with high precision in large-scale environments.

Analyzing Weak Signals

The ability to separate weak warning signals from the usual noise is a complex task and requires that information is interpreted and valued with respect to the context. In order to be able to analyze weak signals, several components are necessary. First of all, a suitable analysis model is needed. A classical approach to address complex problems is to break them down into more manageable sub-problems, solve these separately and then aggregate the results into a solution for the overall problem. This approach is well suited for the analysis of weak signals. The aggregation of sub-problems can be done using several computational approaches such as different kinds of weighted averaging or probabilistic methods. The concept of breaking down a problem into smaller pieces is generally referred to as decomposition. This approach is a suitable solution for gathering, combining, and analyzing information about possible lone actor terrorists.

Figure 1 illustrates a (simplified) example representing a decomposition of the problem of determining whether there is an increased risk that actor X is planning an act of terror. In this example, the original hypothesis (the top node) is broken down to the sub-hypotheses: (1) if actor X has intent to commit an act of terror, (2) if actor X has the capability to commit an act of terror, and (3) if actor X has opportunity to commit an act of terrorism. Each sub-problem can be decomposed further. In this example, the sub-problem intent is decomposed into if actor X is active on radical Internet forums and if actor X is making radical postings. The sub-problem capability is broken down into if actor X is active on Internet forums, is providing postings that reveal certain capabilities, or if actor X is obtaining any materiel that can be used for carrying out terrorist acts.

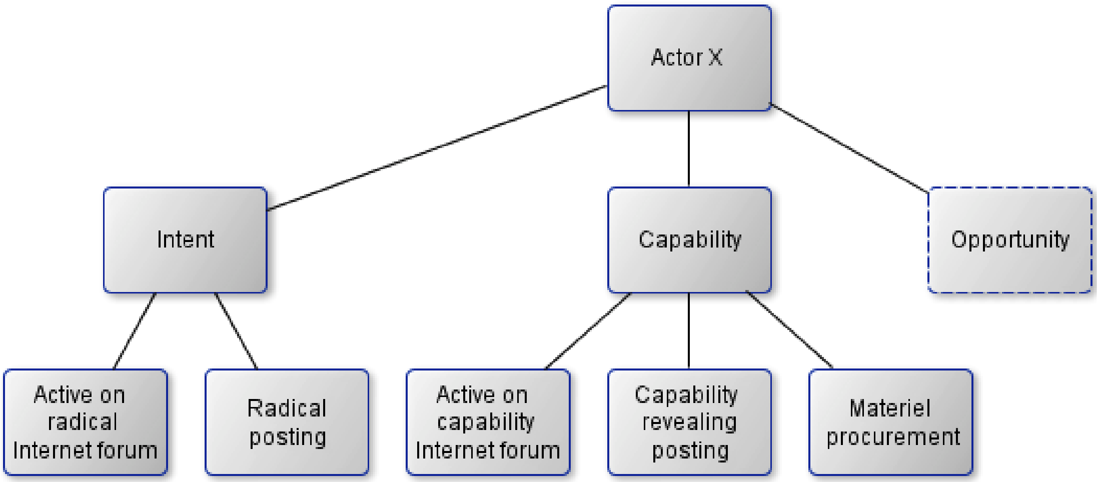
Actor X
Intent
Capability
Opportunity
Active on radical Internet forum
Radical posting
Active on capability Internet forum
Capability revealing posting
Materiel procurement

Figure 1: Decomposition of the overall problem

When the original problem has been broken down to a level that is detailed enough for the current purpose, information that supports each sub-problem can be gathered. The result of each sub-problem can be fused and used to assess whether there is an increased likelihood that someone is planning an act of terrorism. The results can be used to identify potentially dangerous actors that should be subject to further analysis.

By utilizing an analysis model that allows for fusing information from several different sources a more adequate picture of the problem can be provided. The information that can be included in such an analysis model may come from various information sources such as intelligence reports, data from the Internet, medical journals, police reports, and so on.

One example of when the ability to combine information from different sources could have been useful was when Anders Behring Breivik was planning his attacks. Breivik ordered a chemical (aluminum powder) that was listed on the Global Shield[17] list and the Customs Service also had information that he had ordered firing fuse from a company in Poland.[18] Combining this information with the fact that Breivik was a registered owner of three rifles and a gun, and was active on a number of radical discussion forums, could have increased the possibility to detect Breivik before his attack.

Multiple Aliases

A potential problem when trying to fuse various signals collected from social media is that the same individual may use several social media accounts. If an individual uses an account on a discussion forum for expressing radical opinions and a YouTube account for posting videos in which he is firing weapons, this may cause problems for assessing the overall level of threat posed by the individual. For this reason, it can be of interest to use alias matching techniques to find out if several accounts belong to one and the same individual. Such techniques can make use of similarities in user names, similarities in writing style, and similarities in time profiles when trying to match postings written by several aliases.[19] While similarities in user names only can be used if the author is not deliberately attempting to hide that the various accounts belong to the same individual, the stylometric and time-based profiles can be used in most cases. By combining the results from several techniques, surprisingly good accuracy can be achieved with alias matching techniques. So far, most experiments have been made on synthetic datasets in controlled settings, but there are indications that at least stylometric techniques can be used on larger scale in more uncontrolled environments.[20]

[17] Jeffrey T. Wickett, *From Project to a Programme: The Evolution of Global Shield* (World Customs Organization, 2012).

[18] Agrell, *Den svarta svanen och dess motståndare.*

[19] Lisa Kaati, Fredrik Johansson, and Amendra Shrestha, “Detecting Multiple Aliases in Social Media,” Accepted for publication in the proceedings of the 2013 International Symposium on Foundations of Open Source Intelligence and Security Informatics, 2013.

[20] Arvind Narayanan, Hristo Paskov, Neil Gong, John Bethencourt, Emil Stefanov, Richard Shin, and Dawn Song, “On the Feasibility of Internet-Scale Author Identification,” *IEEE Security and Privacy,* 2012.

Targeted Relation Extraction and Automated Profiling

Just because we have the possibility to identify an alias that is active on the Internet and can use various text analysis techniques for getting indications that he or she may be planning an attack, this does not mean that we can say anything about the physical identity of the person behind the alias. In some cases authorities may be able to get information about the IP address that has been used in the communication. However, the IP address may not reveal any information since it is possible to use anonymization techniques such as TOR, computers at public places, or mobile devices with limited possibilities to trace.

One approach to learn more about an Internet user's physical identity is to use what we refer to as targeted relation extraction. The idea is to use natural language processing to detect relations that have an impact on or give clues about someone's physical life. Examples of such relational expressions are: "my wife ...", "I live in ...", "I went to ...", and so on. By extracting such relations, it becomes possible automatically to reason about them and obtain clues that might indicate who the physical person behind the alias is. This kind of techniques can be useful if a larger amount of text is available such as a blog that has been active over a long period of time.

The ability automatically to profile the author of text documents may in theory be useful when we have collected a number of social media posts the real author of which is unknown, for instance, when an alias has been used to generate the posts. There are many examples in the available research literature where researchers try to do various kinds of profiling of individuals based on their writing. Many examples concern classification of the gender or age of the author, but there are also attempts to predict, for instance, the demographics, academic background, cultural background, or psychological profile based on which words that are used, the richness of the used language, word lengths, which syntactical patterns that are most frequently used, and so on. Similar features can also be used for authorship attribution, where one tries to identify the author of a piece of text, given a set of potential authors and their previous known writings. What is common for these kinds of problems is that they tend to be harder to solve the less text material that is available. To separate between various authors based on their writing style is generally much easier when there are just a few potential authors and there is a lot of text available, such as whole books. If there instead is a lot of potential authors and we only have access to a few short blog posts, it becomes much more difficult to discriminate between the various authors.

In theory, one can expect gender to be most easily classified since there are only two classes, male and female. However, although there are only two classes to discriminate between, it is hardly obvious that there is a significant difference in how women generally write, compared to how men write. Some have argued that no difference between male and female writing styles can be expected in many contexts, but results reported by Koppel and others show that algorithms for gender classification are reaching an accuracy of approximately 80 per cent, meaning that around 80 out of 100 randomly selected persons are classified correctly when trying to estimate the gender of the author.[21] The experiments presented by Argamon and others indicate that the most useful style features for gender discrimination are the use of determiners, prepositions, and pronouns.[22] Although it may be useful to know the gender of the author of, say, a series of radical postings, this remains insufficient information to the intelligence analyst. When it comes to classification of age, results presented by Peersman and others suggest that somewhat higher accuracy can be achieved for age than for gender if one only tries to discriminate between adults and adolescents. However, if the number of classes (that is, the number of age groups) is increased, this accuracy can be expected to decrease.[23] In the experiments presented by Argamon and others, the most useful style features for discriminating between age groups are contractions with apostrophes (indicating younger writing), and prepositions and determiners (more often used among older persons).[24]

The use of stylistic text features for determining an author’s native language is studied for texts written in English in Koppel and others.[25] In that paper it is shown that the language patterns used in an author’s native language influence spelling mistakes, function word selection, and grammar usage also in a second language. When trying to classify whether certain English texts have been written by someone having Czech, French, Bulgarian, Russian, or Spanish as mother-tongue, an accuracy of approximately 80 per cent was reported. Once

[21] Moshe Koppel, Shlomo Aragmon, and Anat Shimoni, “Automatically Categorizing Written Texts By Author Gender,” *Literary and Linguistic Computing* 17: 4 (2002), 401-412.

[22] Shlomo Aragmon, Moshe Koppel, James W. Pennebaker, and Jonathan Schler, “Automatically Profiling the Author of an Anonymous Text,” *Communications of the ACM* 52: 2 (2009), 119-123.

[23] Claudia Peersman, Walter Daelemans, and Leona van Vaerenbergh, “Predicting Age and Gender in Online Social Networks,” *Proceedings of the 3rd International Workshop on Search and Mining User-generated Contents* (2011), 37-44.

[24] Aragmon et al., “Automatically Profiling the Author,” 119-123.

[25] Moshe Koppel, Jonathan Schler, and Kfir Zigdon, “Determining an Author’s Native Language by Mining a Text for Errors,” *Proceedings of the eleventh ACM SIGKDD international Conference on Knowledge Discovery in Data Mining* (2005), 624-8.

again, this accuracy can be expected to decrease as the number of potential classes is increased (that is, if adding Swedish and Finnish as possible classes). In general one can expect quite good accuracy when determining to which language family the native language of the author belongs, while it might be harder to decide the exact country from which the author originates.

There are also attempts to classify a writer's personality based on his or her writing. As an example, in Argamon and others, psychology undergraduates were asked to fill in a questionnaire testing for the personality dimensions of neuroticism, extroversion, openness, conscientiousness, and agreeableness. In the next step they were asked to write an essay in twenty minutes concerning their thoughts and feelings. The reported results are significantly better than chance, but still too low to be very usable in practice.[26] Moreover, it can be expected to be even harder to classify an individual's personality from text if he or she is allowed to write about any topic, rather than essays regarding his or her thoughts and feelings.

To summarize, there are many ways in which text analysis techniques can help in determining the author of social media postings, or at least from which category of people the author stems. However, many of the techniques are language-dependent and are often better developed for major languages such as English than for other languages with fewer speakers.

Privacy

Searching and collecting digital traces on the Internet obviously raise privacy concerns. There is nearly always an ongoing debate about privacy issues when it comes to surveillance (whether it is about surveillance cameras, wiretapping, or looking into bank account details data). The surveillance issue usually divides people in two groups: those in favor and those who are against it. Those who are in favor of surveillance usually have arguments such as "if you aren't doing anything wrong, what do you have to hide?" People who are against surveillance argue or reply with comments such as: "If I'm not doing anything wrong, then you have no cause to watch me" or "Because you might do something wrong with my information."

Such concerns were, for instance, voiced when information about the U.S. National Security Agency's (NSA) computer program PRISM (Planning Tool for Resource Integration, Synchronization, and Management) was

[26] Aragmon et al., "Automatically Profiling the Author," 119-123.

leaked to the press in 2013.[27] PRISM had been used for surveillance purposes and had according to the leaked information given access to content held by some of the largest Internet companies. The exposure of PRISM lead to a debate about privacy and the extent to which the U.S. and other governments should be allowed to monitor Internet exchanges.

One could argue that surveillance of any kind is a fine balance between the security of society and the privacy of individuals. Everybody agrees that it is necessary to take measures to prevent and stop terrorist attacks but at what price? Security expert Bruce Schneier argues that "[p]rivacy is an inherent human right, and a requirement for maintaining the human condition with dignity and respect."[28] If surveillance and monitoring of people is excessive, the democratic society that we attempt to protect may lose its value. On the other hand, many would argue that governments cannot stand by and wait until criminal acts are carried out, and they must have the tools and rights to be able to stop attacks before they are carried out.

Just because something is possible or doable, it does not necessarily mean it is a good idea. Social media monitoring is associated with several privacy concerns, and we can expect this debate to continue in the future. We are not attempting to present any solutions to the question of how to find the right balance between privacy and security. Rather, we are here only attempting to present techniques that could be efficient for finding lone actor terrorists. Before such techniques are implemented and put into use, it is vital to investigate all issues regarding privacy versus security carefully to make sure that all aspects of this problem are considered, as well as looking at the problem from a legal and ethical point of view.

Social Media Monitoring and the Future

In this chapter, we have argued that it is common for terrorist organizations to use social media in order to spread information and propaganda. Even though lone actor terrorists by definition are not members of terror networks, we have in previous chapters noted that they become inspired and radicalized through the Internet and social

[27] See, e.g., Barton Gellman and Laura Poitras, "U.S., British Intelligence Mining Data from Nine U.S. Internet Companies in Broad Secret Program," *Washington Post,* 7 June 2013; Glenn Greenwald and Ewen MacAskill, "NSA Prism Program Taps in to User Data of Apple, Google and Others," *The Guardian,* 7 June 2013.

[28] Bruce Schneier, "The Eternal Value of Privacy," www.wired.com/politics/security/commentary/securitymatters/2006/05/70886, accessed on 30 May 2013.

media, and use them for communication with others and for planning their attacks. Tools for monitoring and analyzing the followers of known terrorist organizations on social media and what they are writing about can therefore be an important capability for intelligence agencies whose mission it is to protect society against attacks. The development of natural language processing techniques makes it possible at least in theory to extract various weak signals and warning behaviors such as if someone is planning and preparing an attack, identifies oneself with previous attackers, or expresses very negative or hateful sentiments toward certain groups of people. However, the maturity of automatic processing of text is today insufficient with certainty to separate serious threats from ironic statements or bad jokes, due to the ambiguity of text analysis. Nevertheless, automatic processing may still be very useful to sift through large quantities of text, although human analysts always have to be part of the loop in order to avoid false positives, that is, classifying innocent persons as potential lone actor terrorists. Even though it is likely that natural language processing capabilities will increase further in the future, it is not necessarily the case that social media monitoring is the way to go to fight lone actor terrorism. If such techniques should be used, it is important that they will be designed so as to keep the impact on ordinary citizens’ privacy at a minimum level.