Analysis of Weak Signals for Detecting Lone Wolf Terrorists

Joel Brynielsson, Andreas Horndahl, Fredrik Johansson, Lisa Kaati, Christian Mårtenson, Pontus Svenson
Swedish Defence Research Agency (FOI)
Stockholm, Sweden
Email: firstname.lastname@foi.se

***Abstract*—Lone wolf terrorists pose a large threat to modern society. The current ability to identify and stop these kind of terrorists before they commit a terror act is limited since they are very hard to detect using traditional methods. However, these individuals often make use of Internet to spread their beliefs and opinions, and to obtain information and knowledge to plan an attack. Therefore, there is a good possibility that they leave digital traces in the form of weak signals that can be gathered, fused, and analyzed.**

In this work we present an analysis method that can be used to analyze extremist forums to profile possible lone wolf terrorists. This method is conceptually demonstrated using the FOI Impactorium fusion platform. We also present a number of different technologies that can be used to harvest and analyze information from Internet, serving as weak digital traces that can be fused using the suggested analysis method, in order to discover possible lone wolf terrorists.

***Index Terms*—intelligence analysis; natural language processing; NLP; text mining; affect analysis; weak signals**

I. Introduction

Today, one of the most challenging and unpredictable forms of terrorism are violent terror acts committed by single individuals, often referred to as lone wolf terrorists or lone actor terrorists. These kinds of terror attacks are hard to detect and defend against by traditional police means such as infiltration or wiretapping, since the lone wolves are planning and carrying out the attacks on their own. The problem of lone wolf terrorism is according to many officials presently on the rise and viewed as a greater threat towards society than organized groups. Even though available statistics suggest that lone wolf terrorists accounts for a rather small proportion of all terror incidents [20], they can often have a large impact on the society [8]. Moreover, many of the major terrorist attacks in the United States (with exception for the 2001 attacks against World Trade Center, the Pentagon and the White House) were executed by individuals who were sympathetic to a larger cause—from the Oklahoma City bomber Timothy McVeigh to the Washington area sniper John Allen Muhammad. A similar development can be seen in Europe, where several terrorist attacks have been executed by lone wolf terrorists during the last years. One of the most terrifying acts was the two 2011 terror attacks in Norway committed by Anders Behring Breivik, killing 77 persons in total.

Even though lone wolf terrorists in general cannot be captured by traditional intelligence techniques, this does not imply that there is nothing counterterrorist organizations can do to prevent them. In fact, many lone wolf terrorists are only loners in their offline life, making the Internet an incredibly important source for finding them. According to Sageman [18], most lone wolves are part of online forums, especially those who go on to actually carry out terrorist attacks. The Internet gives isolated lone wolves the opportunity to be a part of a community, something which they often are longing for. There are several communities that encourage and influence individuals to act alone, and individuals that act alone are also often influencing these communities. Online extremist forums and web sites allow for aberrant beliefs or attitudes to be exchanged and reinforced, and creates environments in which otherwise unacceptable views become normalized [21]. In addition to give a possibility of becoming a part of a community, the Internet is also a platform where lone wolves can express their views. The 2010 suicide bomber in Stockholm, Taimour Abdulwahab al-Abdaly, was for example active on Internet and had a YouTube account, a Facebook account and searched for a second wife on Islamic web pages. Anders Behring Breivik used several different social networking sites such as Facebook and Twitter and posted his manifesto "2083, A Declaration of Independence of Europe" on the Internet before committing the two terror attacks in Norway. The actual possession of several social media accounts is obviously perfectly normal, but the content of lone wolf terrorists' social media sites is often far from normal.

One of the major problems with analyzing information from the Internet is that it is huge, making it impossible for analysts to manually search for information and analyze all data concerning radicalization processes and terror plans of possible lone wolf terrorists. In addition to all material that the analysts can find through the use of various search engines, there are also enormous amounts of information in the so called hidden or Deep Web, i.e., the part of Internet that is not indexed by the search engines' web spiders (e.g., due to password protection or dynamically generated content). To produce fully automatic computer tools for finding terror plans is not possible, both due to the large amounts of data and the deep knowledge that is needed to really understand what is discussed or expressed in written text (or other kinds of data available on the Internet, such as videos or images). However, computer-based support tools that aid the analysts in their investigation could enable them to process more data and give better possibilities to analyze and detect the digital

978-0-7695-4782-4/12 $26.00 © 2012 IEEE
DOI 10.1109/EISIC.2012.20

traces [4]. In this paper, we suggest the use of techniques such as hyperlink analysis and natural language processing (including topic recognition and affect analysis) to map the existing dark web forums and to find out which forums and users that can be of interest for human analysts to take a closer look at. In order to combine the outputs from the various suggested methods, we propose using information fusion techniques implemented in FOI's Impactorium fusion platform [22].

The rest of this paper is outlined as follows. In Section II, we give a short background to lone wolf terrorism, and the challenge of finding and identifying such individuals before it is too late. In this paper we mostly focus on weak signals that can be retrieved from the Internet. Therefore, we are in Section III presenting techniques for harvesting digital traces and to analyze these. We propose an analysis method for breaking down the problem of analyzing whether a person is a lone wolf terrorist or not into smaller sub-problems, such as identifying motives (intent), capabilities, and opportunities. These are broken down further, until more concrete indicators are identified that can be fused in order to make an estimate of how probable it is that an individual is a lone wolf terrorist. A discussion on the future potential of this kind of techniques and potential privacy aspects with automatic monitoring and analysis tools is provided in Section IV. Finally, conclusions are presented in Section V.

II. Lone wolf terrorists

The definition of a lone wolf terrorist that will be used throughout this paper is the one used in [6]:

> *A lone wolf terrorist is a person who acts on his or her own without orders from or connections to an organization.*

Lone wolves come from a variety of backgrounds and can have a wide range of motives to their actions. It is observed by [20] that lone wolf terrorists are often creating their own ideologies, combining aversion with religion, society, or politics with a personal frustration. Hence, a lone wolf terrorist can in theory come in any size, any shape, and any ethnicity, as well as representing any ideology [15].

To conduct a successful terror attack, it is necessary to have a number of skills or capabilities. For a lone wolf, obtaining the necessary capabilities for an attack might be a problem, since they can not receive the same kind of systematic training such as, e.g., al-Qaida terrorists. This is one of the reasons why lone wolves rarely are suicide bombers, i.e., such an attack may be too complicated and involves too much preparation [15]. It is also an explanation to why lone wolf terrorists often are using Internet to acquire the knowledge needed to succeed with an attack.

It is not unusual that lone wolf terrorists are sympathizing with extremist movements, but they are not part of or actively supported by these movements. This makes it very hard to discover and capture lone wolf terrorists before they strike, as traditional methods such as wiretapping and infiltration of the organization are not applicable (since there are no networks or organizations to infiltrate). Moreover, it can be very hard to differentiate between those individuals who are really intending to commit an actual terrorism act, and those who have radical beliefs but keep within the law.

A. Digital traces on the Internet

Even though lone wolf terrorists in general are extremely hard to detect by traditional means, there are often many weak signals available that, if detected and fused, can be used as markers of potential interesting behavior that have to be analyzed deeper and investigated further. As has been mentioned by Fredholm [11], nearly all radicalization of lone wolf terrorists take place on the Internet. One example of a well-known online resource inspiring to homegrown terrorism is the online magazine Inspire, published by the organization al-Qaeda in the Arabian Peninsual (AQAP). Internet-based recruitment to terrorist groups is also likely to grow in significance, although recruitment to terror organizations are more often dependent also on offline networks [3], [21], [18]. This kind of Internet-based radicalization processes often result in various digital traces, created when visiting extremist forums, making postings with offensive content, etc. In fact, a notable characteristics of lone wolf terrorists is that they often announce their views and intentions in advance. Once a terror activity has taken place, it is not unusual that e.g., media collect various digital traces in retrospect, and make complains about the police's or intelligence service's ineffectiveness or lack of competence. However, although it can be quite easy to find out the individual pieces once the terror activity already has taken place, it is much more difficult to find out what the relevant pieces are before an actual attack on the society. To find these pieces (i.e., the relevant digital traces), semi-automated analysis is needed since it is impossible for human analysts to manually monitor all the activities of interest on Internet. Such analysis is described in more detail in Section III.

There are a lot of examples of where Internet has been used by lone wolves to spread their views and opinions before an actual attack. One such example is the anti-abortion activist Scott Roeder who killed the physician George Tiller in Kansas, 2009 [2]. Tiller was one of the few doctors in the United States that performed late abortions, and before the attack, Scott Roeader wrote a column on an abortion critical web page where he expressed his views against abortion and Tillers work. Another example of a lone wolf that was using Internet to express his views is James von Brunn, also known as the Holocaust Museum shooter [23]. Von Brunn was an anti-semitic white supremacist who was in charge of an anti-Semitic website where he was able to express his views long before the attack.

III. Techniques for Harvesting and Analyzing Data from the Internet

In this section we present semi-automatic tools and techniques that can be used by intelligence analysts to monitor web sites and forums of interest, analyze the content of these,

and fuse the results in order to discover potential lone wolf terrorists. The goal of the process described in this section is to obtain a list of potential lone wolf terrorists that needs further investigation.

Comparing our suggested approach to related work already described in existing research literature (see e.g., [26], [17], [7]), two main differences can be identified: 1) our focus on lone wolf terrorists rather than terror organizations, and 2) our focus on semi-automated tools for supporting the analyst, rather than fully automated tools.

A. Problem breakdown

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 overarching problem. This approach is well suited for the analysis of weak signals. For each potential threat actor, which in most cases will be represented by one or many aliases (user names), a model is created through the successive decomposition of the threat hypothesis into a number of indicators, corresponding to the weak signals that we want to capture. Figure 1 shows a (simplified) model of how the decomposition of the hypothesis "Actor X is a potential lone wolf terrorist" could look like. At the first level, the hypothesis is separated in three general threat assessment criteria: *Intent* (or *motive*), *Capability*, and *Opportunity*. If all these are met there is a potential risk of an attack. The next level of decomposition shows a number of indicators that can be detected through reconnaissance on the Internet, and one indicator "Materiel procurement" which also could be detected through other information channels.

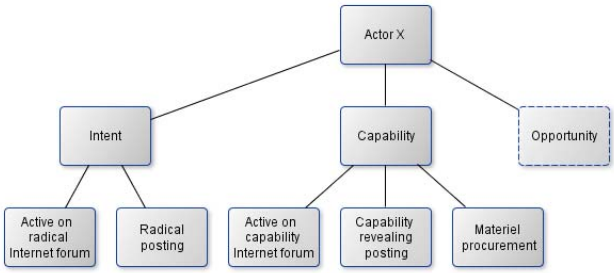

Figure 1. Breakdown of a hypothesis regarding possible lone wolf terrorists.

Once this initial decomposition is done, parallel sub-processes can be started for the various sub-hypotheses. As an example, assuming that an analyst argue that someone needs to have both intent and capability in order to become a lone wolf terrorist, one sub-process can focus on looking for possible motives (e.g., based on radical postings made by the individual) while the other one is focusing on capability (e.g., web sites discussing how to make bombs). The results from the various sub-processes are then fused and used to come up with an estimate of how likely it is that someone is or is starting to become a potential lone wolf terrorist, resulting in a list of potentially dangerous actors to keep an extra eye on.

Each sub-hypothesis represents an *indicator topic*. An indicator topic may consist of several indicators. As an example of how indicator topics are used, consider the motive sub-hypothesis in Figure 1. The sub-hypothesis represents an indicator topic that may have two sub indicators: direct expressions and indirect expressions. Direct expression represents the case when an individual has made a radical statement. Indirect expression represents the case that an individual may have an influential connection to an online message board. The indicators can be assessed based on analyzing online message board activity that shares the same information gathering procedure. How information is collected is described in section B. How to fuse and assess the indicators is discussed in further detail in section C.

The overall process to work with semi-automatic tools using a problem breakdown approach and indicator topics is shown in Figure 2. Initially, the problem is broken down into sub-hypotheses (or sub-problems) and each sub-hypothesis represents an indicator topic. Within the indicator topic the same information gathering procedure can be used. After collecting and fusing information for each indicator topic, alias matching is performed to find authors that uses different aliases. The overall process results in a list of potentially dangerous actors that might be analyzed further.

B. Seeds identification and topic-filtered web harvesting

For each indicator topic identified from the problem breakdown, we propose finding web sites and forums of interest. We will mainly focus on intent here, but a similar methodology can be used for other indicator topics such as capability.

Since the amount of content on the Internet is enormous, it does not make sense to try to search for digital traces from potential lone wolf terrorist without any guidance. Therefore, it is necessary to limit the search and instead focus on a smaller subset of the Internet. Although there are large portions of the Web that is not reachable using search engines such as Google, many extremist web sites are well-known, since part of the idea is to communicate ideologies and other messages to the larger masses. Moreover, a majority of extremist web sites contain links to other extremist sites, according to a study presented in [12]. Hence, it makes sense to use well-known extremist sites as seeds[1], and then try to identify other interesting forums and sites that in some way are connected to the web sites, by using the seeds as a starting point (it is not necessarily so that only extremist web sites are of interest, also "normal" web sites containing information regarding an indicator may be interesting to watch).

The process of systematically collecting web pages is often referred to as crawling. Usually, the crawling process starts from a given source web page (the seeds described above) and follows the source page hyperlinks to find more web pages [13]. The crawling process is repeated on each new page and continues until no more new pages are discovered

[1]The actual seeds to use are up to the analyst to define and are outside the scope of this paper.

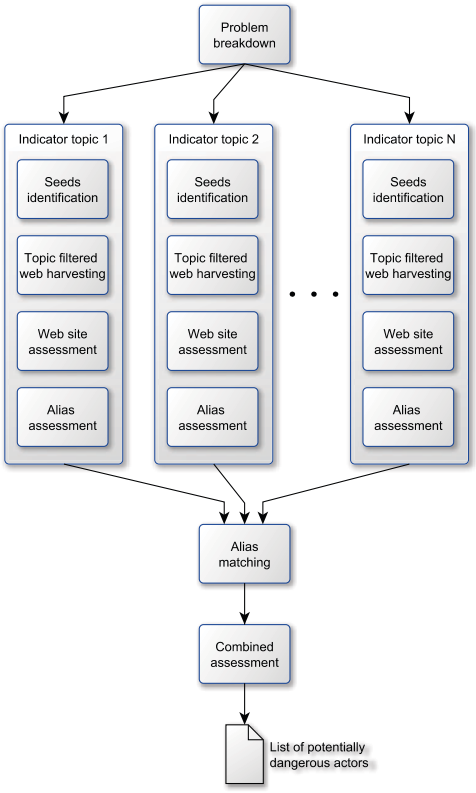

Figure 2. The overall process using a problem breakdown approach and indicator topics.

or until a certain number of pages (that have been determined beforehand) have been collected. By treating the collected web sites as nodes in a graph, and by creating an edge between two web sites each time a hyperlink is found between them, it becomes possible to create a (large) network that can be analyzed further to find out which the most interesting web sites are. By using hyperlink analysis a large number of potential extremist forums can be found. However, many of the web sites will be perfectly normal, making them rather uninteresting for intelligence analysts. Hence, it is of uttermost interest to be able to automatically separate web sites with an interesting content from the ones with normal, uninteresting content (that is, from a counterterrorist perspective). In order to make this kind of analysis, natural language processing (NLP) and text mining can be of great use. We suggest having a predefined list of keywords to search for on the crawled web pages. If enough of the terms are encountered on a web page, it is marked as interesting and the web site is added to the queue. However, if they are marked as irrelevant, the web page becomes discarded, and no links are followed from it. The same holds true for URLs that are part of a *white list*, to which the analyst can choose to add web sites matching the keywords but are judged not to be relevant for further analysis (e.g., web sites with the purpose of countering extremist propaganda). While crawling the web it is also possible to discard links that are broken. If a web site is inaccessible due to password protection, the analyst can be asked to either choose to discard the link, or to manually create a user login and enter the user credentials to access material on the site. Our suggested approach is in many ways similar to the approach used for identifying online child pornography networks in [14].

To evaluate our web mining approach, we have implemented a proof-of-concept web spider. The goal is to create a network consisting of web sites, forums (message boards, discussion boards), forum posts and aliases. An example of such a network can be found in Figure 3. As can be noted, the network becomes very large and therefore it is important to prune the network using natural language techniques. The spider is based on the crawler Crawler4J[2] and extended with methods for Internet forum information extraction.

Given a set of seeds (web page URLs), the web spider expands the network by following all links that can be found in the page that meet a set of conditions. First of all the link should point to a web page, and secondly the content of the web page should be classified as interesting (matching a list of one or several predefined keywords). If the page represents a discussion forum, tailored content extraction algorithms are applied. The algorithms extract the user aliases and their posts, and add this information to the network (to be further used in the web site and alias assessment phases). In our initial proof-of-concept implementation, we have developed information extraction algorithms for a specific representative Internet forum.

In a real-world setting, one need to address the fact that Internet forums or web sites may have significantly different structures. Hence, a flexible strategy for learning the structure of a new site is desirable. One way to overcome this obstacle is to let an algorithm guess the structure, try to extract relevant information and let a human (the analyst) verify the results. Another way is to let humans analyze the HTML representation and locate specific HTML tags that can be used as markers for where to find relevant information and how to separate posts.

C. Web site and alias assessment

Once a list of interesting web sites or forums have been created using topic-filtered web harvesting, the idea is to make a deeper analysis of postings on these, by making use of natural language processing and text mining techniques. One type of text mining known as affect analysis has earlier been identified as being useful for measuring the presence of hate and violence in extremist forums [1]. To be able to use natural language processing techniques, it is necessary to first preprocess the retrieved content from the web sites. This pre-processing for example includes removing HTML tags, and tokenizing the text into sentences. The sentences are then fed

[2]Downloadable from: http://code.google.com/p/crawler4j/

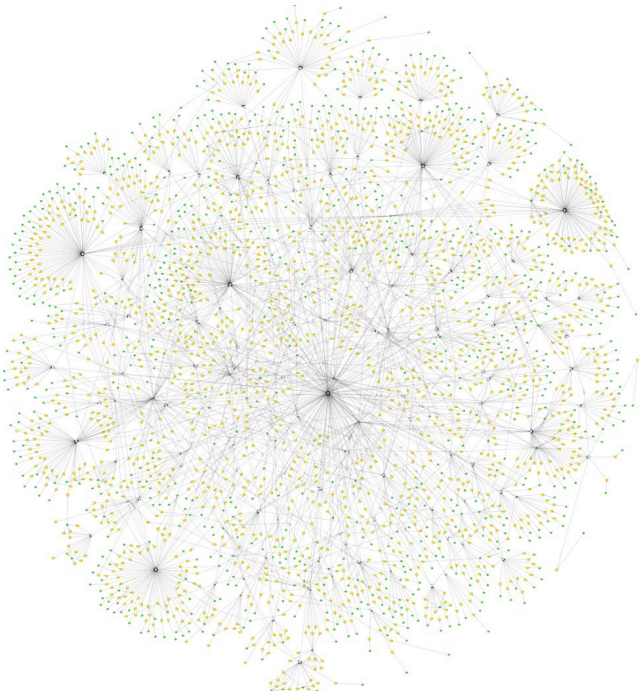

Figure 3. A network graph created by our web spider based on a single seed. Nodes in the network represent discussion boards, posts and aliases.

to a classifier that suggests a level of interestingness for the web site (in the case of intent/motive, the interestingness may refer to level of radicalization, hate, etc.). This estimate is used to create a list of the most interesting sites at the time being, as well as being used as input to the calculation of a user's level of interestingness (e.g., in terms of radicalization level), as explained below.

Classifiers for estimating the level of radical content or other types of interestingness in a text (e.g., a blog post or a tweet) can be built in various ways. One alternative is to manually create a discriminant-word lexicon that can be used for classifying the text; the higher fraction of terms in the text present in the lexicon, the higher the level of interestingness. To manually create such a list may, however, be a tricky task, and it may also be necessary to update the list with regular intervals, as the popular words to express radical opinions or other kinds of topics may change over time. Within the research field of text mining, it has been shown that handcrafted lexicons are often not the best alternative for text classification tasks. Instead, various unsupervised and supervised learning algorithms are more frequently used. Irrespectively of which type of technique that is used, some input will be needed from an expert. In case a handcrafted list of words is used, the actual terms to use have to be specified by experts. In the case of an unsupervised approach, a list of seed terms has to be suggested by the experts which then can be used to automatically find and classify other terms that, e.g., are synonyms or antonyms to the manually labeled terms, or in other ways are co-occuring with terms with a known label. Finally, in the supervised case, the expert has to manually classify a number of text samples into the classes $radical$ and $non\text{-}radical$ (or $interesting$ and $non{-}interesting$ in the more general case). It can be expected that the supervised approach will yield the best performance, but this comes with a cost of finding useful data for training purposes, and the manual annotation of the training data.

One type of classifier that often is used for various supervised natural language classification tasks is the naïve Bayes classifier. An advantage of such an approach is that it is easy to interpret for humans, making it possible to verify that a learned model looks reasonable. Furthermore, it is more computationally effective than many alternative algorithms, making the learning phase faster. In order to use such as classifier for discriminating between texts with $radical$ and $non\text{-}radical$ content, a natural first step would be to tokenize the text. By extracting features such as unigrams (single words), bigrams (pairs of words) or trigrams (triples of words) from the tokenized text, this can be used for training the classifier and to classify new texts once the classifier has been trained. Since there would be very many features if allowing for all possible unigrams and bigrams, a necessary step would be feature reduction, in which the most informative features $f_1, \ldots, f_n$ are selected from the training data and used as leaf nodes in the resulting classifier. By extracting features from new texts to be classified, we can according to Bayes' theorem calculate the posterior probability of the text having a certain label (e.g., $radical$ or $non\text{-}radical$) as:

$$P(label|f_1, \ldots, f_n) = \frac{P(label)P(f_1, \ldots, f_n|label)}{P(f_1, \ldots, f_n)}. \quad (1)$$

Now, by using the condional independence assumption of the naïve Bayes model, this is reduced to:

$$P(label|f_1, \ldots, f_n) \propto P(label) \prod_{i=1}^{n} P(f_i|label). \quad (2)$$

This conditional independence assumption is rather strong and does not necessarily hold in practice. Given the class label, the occurrence of a word is not independent of all other words, even though this is assumed in Equation 2. This may result in that conditionally dependent words can have too much influence on the classification. Despite this, naïve Bayes have been shown to work well for many real-world problems.

The needed probabilities on the right side of Equation 2 can easily be estimated from the training data (using Laplace smoothing to account for zero counts).

Other popular choices for text classification tasks is the use of maximum entropy classifiers (relying on the principle of choosing the most uniform distribution satisfying the constraints given by the training data) or support vector machines. Irrespectively of what choice of classifier that is made, the most important part is to get hold of enough training data of good quality. Once this is solved, the next big question is which features to use. To use unigrams as features is the most straightforward way and will most likely be enough to separate terrorism-related discussions from many other kinds

of discussions of no relevance to the subject. However, it is not obvious that unigrams are enough for more fine-grained classification, e.g., separating between postings where terrorist acts are discussed or reported on, and where intentions to actually commit terrorism acts are expressed. It may therefore be beneficial to use bigrams or trigrams to allow for a less shallow analysis.

1) Ranking of aliases: Figure 4 illustrates how each alias is connected to a set of posts and the discussion boards where the posts were made. The fact that an author made a post on a discussion board indicates that the author is interested in the topics discussed in the message board and therefore may be influenced by other posts on the specific discussion board.

We formalize this dependence in the following way. From the set of web sites $\mathbb{W}$ that where considered to be interesting according to the web site assessment, a list $\mathbb{X}$ of all present aliases is extracted. For each alias $x \in \mathbb{X}$, we compute the interestingness based on the information gathered about x. The interestingness function is a function

$$I : \mathbb{X} \to \Gamma, \tag{3}$$

where Γ is a suitable representation of rankings e.g., $\Gamma = \{low, medium, high\}$ or $\Gamma = \mathcal{N}$ where $\mathcal{N}$ is the set of natural numbers. Similarly we define an interestingness function for a web site

$$J : \mathbb{W} \to \Gamma, \tag{4}$$

where $\mathbb{W}$ is the set of web sites or discussion boards and Γ is a suitable representation of rankings. The interestingness function for a discussion board can be computed using natural language processing and text mining techniques or manual analysis performed by a analyst or a combination of these.

We write the interesting function $I(x)$ as follows:

$$I(x) = f(\{J(w) : w \in \mathbb{W} \wedge A(x, w) = 1\}, C(x)), \tag{5}$$

where $A(x, w)$ is equal to 1 if x is active (has posted something) on discussion board $w \in \mathbb{W}$ and 0 otherwise. $C(x)$ is a content analysis function which represents the content that x has posted on all discussion boards considered. The content analysis function could for example be a measurement of the number of radical words that x has used.

The fusion function f needs to be adapted to the specific problem since there are many aspects and dependences should be considered. We will investigate this function further in future work.

D. Alias matching and fusion of the obtained indicators

Alias matching on a specific site or forum is often trivial as long as an individual is not using many different aliases. However, if a user is active on a number of web sites, forums, or other kinds of social media, alias matching can be very cumbersome. If people are using the same alias everywhere, it is simple, and if there are only small variations in user names, entity matching approaches such as the Jaro-Winkler distance metric [25] may be useful. However, if people are using aliases which are more or less arbitrarily selected, the actual user name as such cannot be used for the matching process. If messages have been posted on non-radical forums it might be possible for police or intelligence services to get information about the IP address that has been used when making the posting, but this cannot be expected to be retrieved from extremist forums. Moreover, the IP address may not necessarily be of interest, since people can use dynamic IP numbers, use computers at Internet cafes, etc. As an alternative, it could be of interest to look into using author recognition techniques. The idea with author recognition is to determine who wrote a text by studying specific characteristics of the text. Such characteristics could for example be choice of words, language, syntactic features, syntactical patterns, choice of subject, or different combinations of these characteristics [16]. Author recognition is a difficult problem, especially on short texts such as posts on discussion boards. In [9], it is noted that using characteristics such as smileys gives a better result when recognizing authors from short texts. One direction for future work would be to investigate further to what extent adding different characteristics that are specific for writing on discussion boards can be used to improve the results of author recognition.

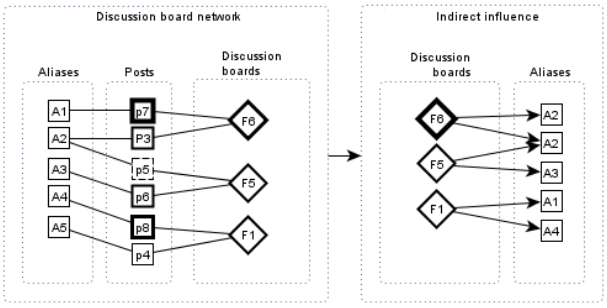

Figure 4. The relations of aliases, posts and discussion boards. A directed link between a message board and an alias represents the fact that an actor (behind an alias) may be influenced by the content of other posts. The thickness of the node lines reflects the level of interestingness of the post or discussion board.

Social network analysis (SNA) [19], [24] could also be used to help in the identification of authors by computing structural similarities between different aliases. If two aliases post to the same forums, on the same topics, and regularly comment on the same type of posts, it is likely that they are in fact the same. It is also possible to use abstraction techniques such as simulation [5] to determine the likelihood with which two aliases are the same.

The values of the different indicators are fused in order to come up with an answer to the original problem, i.e., to which degree the collected evidence or weak signals support the hypothesis that an individual is (or will become) a lone wolf terrorist. The fusion can be made using the Impactorium tool [10], which is used to create top-down threat models as the one presented earlier. A screen shot of how the values of a threat model are inferred in the Impactorium tool is shown

in Figure 5.

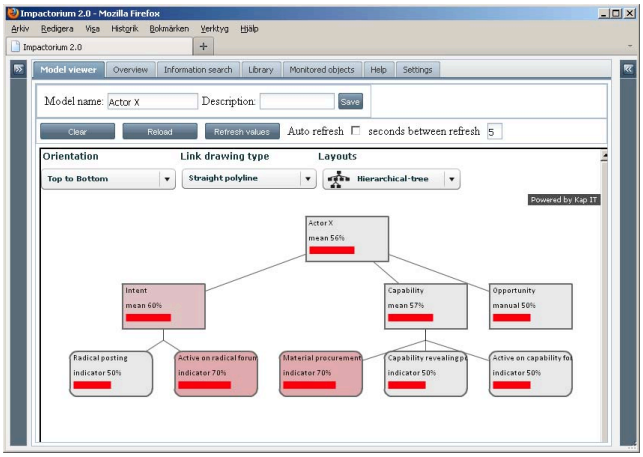

Figure 5. A threat model in the Impactorium tool, where a number of evidences have been fused.

Various combination functions such as min, max, average or weighted sum can be used to deduce and reason whether a sub-hypothesis or the initial hypothesis is likely to occur. Except for combining the various digital traces that have been collected, Impactorium also allows for fusion with information coming from other sources, such as intelligence reports or data from sensors. As an example, if customs provide information that an individual has bought large quantities of fertilizers, this information can be inserted into the threat model calculations.

For each alias that is considered interesting, a threat model is created and information is gathered. Based on the results of the fusion, a list of aliases worth monitoring more closely is created. An example of such a list is shown in Figure 6.

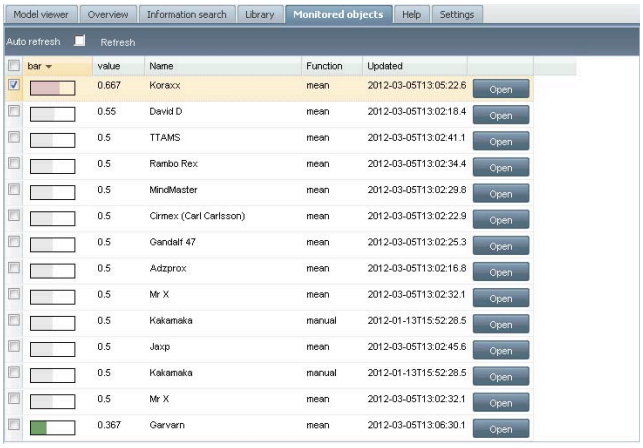

Figure 6. List of monitored aliases within the Impactorium tool

Since the content of web sites such as extremist forums is not static, this is a process that has to be done over and over again, the first stages can however be done more seldom than the later phases, since forums and web sites of interest will pop up or become obsolete on a much slower rate than the change in content of the web sites. It is also important to note that duration of time is a significant factor in this process. It is very likely that becoming a lone wolf terrorist is not something that happens over night, instead this process can take several years.

IV. Discussion

The search for digital traces on Internet that can be fused in order to try to find potential lone wolf terrorists is a fine balance between people's security at the one hand, and people's privacy on the other. To automatically search through large masses of text and use text mining techniques to try to identify whether a text should be treated as radical or not can by some people be seen as a violation of privacy. The needs of the law enforcement and intelligence communities and the right to privacy must be balanced. It should however be noted that analysts are checking extremist forums already today. It is always a human analyst that should check the reasons for why a user has been classified as having a potential motive or intent of being a potential lone wolf terrorist, and if actions should be taken to bind an alias to a physical person, and to collect more information using other means. The analyst can also always decide on whether an alias should be removed from the list of "suspect" individuals. This highlights the need for a mixed initiative system with a human-in-the-loop as a central component.

Having such a human-in-the-loop makes is possible to tolerate a higher number of false positives than would be possible in a fully automated system. Since there is a trade-off between false positives and false negatives, the increase of false positives should decrease the number of false negatives (i.e., classifying weak signals from potential terrorists as non-interesting). The suggested method should be thought of as a help for the analyst to filter out a smaller set of data to look at, rather than a method to be fully automated.

While we here have focused on text, it is worth noticing that a lot of material posted to web sites and social media is not text. On extremist forums, it is not unusual with video clips of executions, bomb making instructions, etc. There is a lot of ongoing research on image and video-content analysis, as well as content-based image retrieval (CBIR) that can be useful in the future, but as far as we know, no mature techniques for identifying radical content in video with good precision exists today. Another possibility is to automatically extract speech from audio and video content and transcribe it into text. Such technology is, e.g., available in a beta version for certain English-language videos on YouTube. The technology is still far from perfect, but it can be expected that it will work well in the foreseeable future, and then also for other languages than English.

The techniques we have proposed in this paper are not constrained to work for a single language. Many resources for text mining (such as WordNet) are language dependent and only works for English, but the natural language processing techniques we have suggested are not relying on such resources. However, it is not possible to construct a

discriminant-word lexicon for, e.g., Swedish, and expect it to work for web sites written in Russian or Arabic. One way to deal with content in several languages is to develop separate lexicons for the various languages of interest. Another way that demands less resources is to preprocess the text by automatic machine translation into a common language, and then use the preprocessed text as input to the classifier. Such an approach will probably give worse precision, but will demand less resources.

V. Conclusions

One of the major problems with detecting possible lone wolf terrorists is that there is no consistent or typical profile of a lone wolf. Moreover, the lone wolves are hard to capture using traditional intelligence methods since there are no physical groups to infiltrate or wiretap. However, there are many concrete actions and activities taken by an individual (that are not necessarily illegal) that can be treated as weak signals and that combined may indicate an interest in terrorism acts. Recognizing and analyzing digital traces from online activites of possible lone wolf terrorists is one aspect in the difficult problem of detecting lone wolf terrorists before they strike. We have presented a framework for working with such digital traces through the use of techniques such as topic-filtered web harvesting and content analysis using natural language processing. Parts of the proposed system have been implemented, while work remains to be done for other parts.

It is important to highlight that the proposed system is not intended to be fully automatic. The central component of the system will be the human analyst, but this analyst will be supported in the work of finding, analyzing, and fusing digital traces of interest for finding potential lone wolf terrorists.

Acknowledgements

This research was financially supported by Vinnova through the Vinnmer-programme, and by the the FOI research project Tools for information management and analysis, which is funded by the R&D program of the Swedish Armed Forces.

References

[1] Ahmed Abbasi and Hsinchun Chen. Affect intensity analysis of dark web forums. In *Proceedings of the 5th IEEE International Conference on Intelligence and Security Informatics*, 2007.

[2] Robin Abcarian. Scott Roeder convicted of murdering abortion doctor George Tiller. In *Los Angeles Times*, January 29, 2010.

[3] Anthony Bergin, Sulastri Bte Osman, Carl Ungerer, and Nur Azlin Mohamed Yasin. Countering internet radicalisation in southeast Asia. Technical Report 22, ASPI, March 2009.

[4] Joel Brynielsson, Andreas Horndahl, Lisa Kaati, Christian Mårtenson, and Pontus Svenson. Development of computerized support tools for intelligence work. In *Proceedings of ICCRTS 2009*, 2009.

[5] Joel Brynielsson, Lisa Kaati, and Pontus Svenson. Social positions and simulation relations. *Journal of Social Network Analysis and Mining*, 2(1):39–52, 2012.

[6] Fred Burton and Scott Stewart. The lone wolf disconnect. Terrorism Intelligence Report - STRATFOR, 2008.

[7] Hsinchun Chen, Edna Reid, Joshua Sinai, Andrew Silke, and Boaz Ganor, editors. *Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security*, volume 18 of *Integrated Series in Information Systems*. 2008.

[8] COT. Lone-wolf terrorism. Technical report, Instituut voor Veiligheids- en Crisismanagement, 2007.

[9] Marcia Fissette. Author identification in short texts. Master's thesis, Radboud University Nijmegen, 2010.

[10] Robert Forsgren, Lisa Kaati, Christian Mårtenson, Pontus Svenson, and Edward Tjörnhammar. An overview of the Impactorium tools. In *Proceedings of SWIFT2008*, 2008.

[11] Michael Fredholm. Hunting lone wolves - finding islamist lone actors before they strike. In *Stockholm Seminar on Lone Wolf Terrorism*, 2011.

[12] Phyllis B. Gerstenfeld, Diana R. Grant, and Chau-Pu Chiang. Hate online: A content analysis of extremist internet sites. *Analyses of Social Issues and Public Policy*, 3(1):29–44, 2003.

[13] Monika R. Henzinger. Hyperlink analysis for the web. *IEEE Internet Computing*, 5:45–50, 2001.

[14] Kila Joffres, Martin Bouchard, Richard Frank, and Bryce Westlake. Strategies to disrupt online child pornography networks. In *Proceedings of the 2011 European Intelligence and Security Informatics Conference*, pages 163–170, 2011.

[15] Lisa Kaati and Pontus Svenson. Analysis of competing hypothesis for investigating lone wolf terrorists. In *Proceedings of the 2011 International Symposium on Open Source Intelligence and Web Mining*, 2011.

[16] Sangkyum Kim, Hyungsul Kim, Tim Weninger, and Jiawei Han. Authorship classification: a syntactic tree mining approach. In *Proceedings of the ACM SIGKDD Workshop on Useful Patterns*, 2010.

[17] Edna Reid, Jialun Qin, Wingyan Chung, Jennifer Xu, Yilu Zhou, Rob Schumaker, Marc Sageman, and Hsinchun Chen. Terrorism knowledge discovery project: a knowledge discovery approach to addressing the threats of terrorism. In *Proceedings of the Second Symposium on Intelligence and Security Informatics, ISI 2004*, 2004.

[18] Marc Sageman. *Leaderless Jihad: Terror Networks in the Twenty-First Century*. University of Pennsylvania Press, 2008.

[19] John Scott. *Social Network Analysis: A Handbook*. Sage Publications, London, 2 edition, 2000.

[20] Ramón Spaaij. The enigma of lone wolf terrorism: An assessment. *Studies in Conflict & Terrorism*, 33(9):854–870, 2010.

[21] Tim Stevens and Peter R. Neumann. Countering online radicalisation: A strategy for action. Technical report, International Centre for the Study of Radicalisation and Political Violence, 2009.

[22] Pontus Svenson, Robert Forsgren, Birgitta Kylesten, Peter Berggren, Wong Rong Fah, Magdalene Selina Choo, and J.K.Y. Hann. Swedish-singapore studies of bayesian modelling techniques for tactical intelligence analysis. In *Proccedings of the 13th International Conference on Information Fusion*, 2010.

[23] James von Brunn. An ADL backgrounder beliefs and activities. In *Anti-Defamation League*, June 11, 2009.

[24] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[25] William E. Winkler. String comparator metrics and enhanced decision rules in the fellegi-sunter model of record linkage. In *Proceedings of the Section on Survey Research Methods*, pages 354–359, 1990.

[26] Li Yang, Fiqiong Liu, Joseph M. Kizza, and Raimund K. Ege. Discovering topics from dark websites. In *Proceedings of the IEEE Symposium on Computational Intelligence in Cyber Security (CICS2009)*, 2009.