2017-12

Software requirement specifications for a social-media threat assessment tool

Barnett, Craig T.

Monterey, California: Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

SOFTWARE REQUIREMENT SPECIFICATIONS FOR A SOCIAL-MEDIA THREAT ASSESSMENT TOOL

by

Craig T. Barnett

December 2017

Thesis Advisor:	Rodrigo Nieto-Gomez
Second Reader:	Lauren Wollman

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704–0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 2017

3. REPORT TYPE AND DATES COVERED
Master's thesis

4. TITLE AND SUBTITLE
SOFTWARE REQUIREMENT SPECIFICATIONS FOR A SOCIAL-MEDIA THREAT ASSESSMENT TOOL

5. FUNDING NUMBERS

6. AUTHOR(S) Craig T. Barnett

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB number ____N/A____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

Police officers are often the targets of threats, both verbal and written. Twitter and Facebook allow the communications of these threats quickly, anonymously and in high volume. Law enforcement agencies become overwhelmed trying to determine which are the most serious, since they have limited investigators. Identifying threats that have a high likelihood of violence is also very subjective. How can risk assessment of these threats be improved? As an answer to this question, a software-requirement specification document details a new software that starts the threat assessment process earlier. This software incorporates a social media and language sentiment analyzer, criminal history information and threshold, and confidence scoring to alert law enforcement of threats likely to end in violence. Twitter and Facebook posts that reach a predetermined score alert investigators of a high probability threat on which investigators can focus their efforts. During the development of the software proposal, this thesis finds that implementing this software could improve law enforcement intervention to threats communicated over social media.

14. SUBJECT TERMS
law enforcement, threats, social media, confidence scoring, software, software requirement specifications, sentiment analyzer, Twitter, Facebook

15. NUMBER OF PAGES
83

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

THIS PAGE INTENTIONALLY LEFT BLANK

SOFTWARE REQUIREMENT SPECIFICATIONS FOR A SOCIAL-MEDIA THREAT ASSESSMENT TOOL

Craig T. Barnett
Lieutenant, Raleigh (North Carolina) Police Department
B.S., Florida State University, 1992

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF ARTS IN SECURITY STUDIES
(HOMELAND SECURITY AND DEFENSE)

from the

NAVAL POSTGRADUATE SCHOOL
December 2017

Approved by: Rodrigo Nieto-Gomez
Thesis Advisor

Lauren Wollman
Second Reader

Eric Dahl
Associate Chair of Instruction,
Department of National Security Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Police officers are often the targets of threats, both verbal and written. Twitter and Facebook allow the communications of these threats quickly, anonymously and in high volume. Law enforcement agencies become overwhelmed trying to determine which are the most serious, since they have limited investigators. Identifying threats that have a high likelihood of violence is also very subjective. How can risk assessment of these threats be improved? As an answer to this question, a software-requirement specification document details a new software that starts the threat assessment process earlier. This software incorporates a social media and language sentiment analyzer, criminal history information and threshold, and confidence scoring to alert law enforcement of threats likely to end in violence. Twitter and Facebook posts that reach a predetermined score alert investigators of a high probability threat on which investigators can focus their efforts. During the development of the software proposal, this thesis finds that implementing this software could improve law enforcement intervention to threats communicated over social media.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

API	application program interface
ATAP	Association of Threat Assessment Professionals
CAD	computer aided dispatch
IEEE	Institute of Electrical and Electronic Engineers
IP	internet protocol
RMS	report writing management software
SRS	software requirement specification
SQL	structured query language

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Police officers are often the subjects of threats. Investigations of these threat cases follow a framework similar to one published in 2006 by the Association of Threat Assessment Professionals (ATAP).[1] The framework provides behavioral indicators and risk factors gleaned from information of a known suspect, which is analyzed and applied during threat management activities.[2] Nevertheless, since the publication of that framework in 2006, the volume of threats to law enforcement personnel through social media—notably Facebook and Twitter—has increased. Whereas people once handwrote threats, they now simply send messages from their phones or computers. The ease and instant connectivity of social media means a much higher volume of threats than before, which taxes law enforcement's ability to investigate each one.

Complicating the investigation is the fact that not every threat is a real one. Calhoun and Weston divide threateners into two categories, hunters and howlers, demonstrating that not all threats lead to violent acts.[3] The internet magnifies the ability of people to communicate threats, but very few are carried out. Determining which threats are real is difficult. First of all, Twitter and Facebook provide a platform of communications that can hide someone's identity. Obtaining the identity of a Twitter account holder requires court paperwork based on probable cause or exigent circumstances. Second, social media allow people to repost tweets from the original threatener, whereby the same threat appears to originate from many different people. In this case, it is difficult to identify which poster poses the highest risk of violence.

Analyzing the language of Twitter and Facebook posts could provide an earlier starting point. Patton et al. illustrate how language from Twitter messages can be coded

[1] Association of Threat Assessment Professionals, *Risk Assessment Guideline Elements for Violence: Considerations for Assessing the Risk of Future Violent Behavior* (Sacramento, CA: Author, 2006), https://c.ymcdn.com/sites/www.atapworldwide.org/resource/resmgr/imported/documents/RAGE-V.pdf.

[2] Association of Threat Assessment Professionals, 5.

[3] Frederick S. Calhoun, and Stephen W. Weston, *Concepts and Case Studies in Threat Management* (Boca Raton: CRC Press, 2013), 11, http://www.crcnetbase.com.libproxy.nps.edu/isbn/9781439892183.

for different types of aggression—direct, indirect, proactive, and reactive.[4] The coding method is not used to assess risk. Coding Twitter messages also occurs in the private sector. Automated analysis of social-media language can identify risks to a company's reputation through a coding process.[5] By coding text, analysts can divide tweets into negative or positive feelings about the company.[6] Progressing even further, confidence scoring provides a ranking based on how probable an event is to happen.[7] Using this tool could also help determine which threats have a higher probability of happening. Lastly, tools such as public record databases, search engines and social media analysis software exist which could analyze information provided in threatening posts, feed information to the confidence scoring tool and possibly increase the accuracy of confidence.

In sum, risk assessments begin at the point a suspect is identified. The evolution of social networking sites allows individuals to communicate threats anonymously and in high volume. This ability means investigators cannot wait to identify a suspect to begin a risk assessment. There are software tools already available that can help start assessments earlier. By creating a new software platform that combines a social media monitoring tool, a language sentiment tool, a criminal history database and a confidence scoring tool, law enforcement identifies violent people before they injure or kill their victims. Although more development is needed, the case stories in this thesis shows the proposed software correctly identifying people that post on social media and then act out violently.

[4] Desmond U. Patton et al., "Gang Violence on the Digital Street: Case Study of a South Side Chicago Gang Member's Twitter Communication," *New Media & Society* (January 2016): 7, doi: 1461444815625949.

[5] Paul Alpar, and Daniel Ohliger, "Creation of Risk Profiles of Business Customers from Social Media," *Banking and Information Technology* 16, no. 1 (March 2015): 26, http://web.a.ebscohost.com.libproxy.nps.edu/ehost/pdfviewer/pdfviewer?sid=a77475c8-66bb-44f6-9007-565450b5762d%40sessionmgr4007&vid=1&hid=4214.

[6] Alpar and Ohliger, 26.

[7] Bill Murdock, "How to select a threshold for Acting Using Confidence Scores," IBM Watson, June 23, 2016,https://developer.ibm.com/watson/blog/2016/06/23/how-to-select-a-threshold-for-acting-using-confidence-scores/.

ACKNOWLEDGMENTS

I would like to first thank my family for their support during not only this degree but also the entirety of my career. They have endured my missing many events and weekends for my "day job" and have now supported me while away for 2-week stretches and then weekends of schoolwork. To my wife Blanka, daughter Alex, and son Eli, I thank you for always supporting me.

Secondly, I want to thank the Raleigh Police Department and Chief Cassandra Deck-Brown for allowing me the opportunity to expand my education. The support, encouragement and interest in my education from the department as a whole relieved a lot of the stress during the process.

Finally, to my thesis committee, I say thank you. Rodrigo Nieto-Gomez's help in shaping my idea and support where my knowledge of technology was not enough to keep up allowed me to get to the end. Lauren Wollman, as my reader, helped ensure that I presented a final product that is not only academic but also functional. As a true member of my thesis committee, I also want to thank Noel Yucuis. Her help as a research and writing coach was invaluable to my project. The time spent going over her "red ink" made me see I actually can put words onto paper.

THIS PAGE INTENTIONALLY LEFT BLANK

I. THREAT INVESTIGATIONS NEED TO CHANGE

A. RESEARCH QUESTION

This thesis answers the following question: How can we improve risk assessment used by law enforcement on threats communicated over Twitter and Facebook?

B. PROBLEM STATEMENT

> *February 29, 2016*
>
> *The day began like any other day for us at the Raleigh Intelligence Center. Normal for us is juggling the constant influx of requests from various district captains and officers while we also complete our regular crime analysis products. The craziness came to a sudden stop as we heard an officer on the radio call out a foot pursuit, followed by silence. When back-up finally arrived to help the officer, radio traffic conveyed that the suspect had been killed by the officer during a struggle over a gun. Soon, the news monitors showed an increasingly tense scene: the streets filled with people pushing the crime-scene tape as far in as they could before it broke, yelling and pointing fingers at the cops who stood stoically and expressionless behind the tape.*
>
> *Before our eyes, the events of February 29 unfolded on network news, and push notifications from social media flooded our detectives' monitors. Live video from the scene flowed in, and as the investigation wrapped up, protest groups began streaming videos and posting messages over Twitter and Facebook. Later that day, social media advertised vigils and meetings to organize civil protests. Some of these messages spewed hatred at the police.*
>
> *The next day, the department released the name of the officer. Posts filled with anger turned into threatening statements directed at the police. One Twitter post read, "Fuck Twiddy [the officer]. Sumbody needa kill his wife nd kids, make his ass feel it!!!" Comments and reposts of these tweets gave them a life of their own. If you printed off every social media page that contained a threat or a repost of a threat, we would have had a stack three feet tall. Tasked with assessing threats by over 60 different people, our team had to find a new way. We didn't have enough investigators to investigate each one immediately. We weren't even sure whether they were all real, especially since some people had simply reposted others' threats. How could we sort through the posts and identify the highest risk posts?*

Police officers are often the subjects of threats. Investigations of these threat cases follow a framework similar to one published in 2006 by the Association of Threat Assessment Professionals (ATAP).[1] The framework provides behavioral indicators and risk factors gleaned from information of a known suspect, which is analyzed and applied during threat management activities.[2] Nevertheless, since the publication of that framework in 2006, the volume of threats to law enforcement personnel through social media—notably Facebook and Twitter—has increased. Whereas people once handwrote threats, they now simply send messages from their phones or computers. The ease and instant connectivity of social media means a much higher number of threats than before, which taxes law enforcement's ability to investigate each one.

Complicating the investigation is the fact that not every threat is a real one. Calhoun and Weston divide threateners into two categories, hunters and howlers, demonstrating that not all threats lead to violent acts.[3] The internet magnifies the ability of people to communicate threats, but very few threats are carried out. Determining which threats are real is difficult. First of all, Twitter and Facebook provide a platform of communications that can hide someone's identity. Obtaining the identity of a Twitter account holder requires court paperwork based on probable cause or exigent circumstances. Second, social media allow people to repost tweets from the original threatener, whereby the same threat appears to originate from many different people. In this case, it is difficult to identify which poster poses the highest risk of violence.

Assessing the risk of violence begins at the point a suspect is identified.[4] An aspect of typical risk-assessment protocols suggests interviewing the threatening subject. Borum et al. provide an example of these protocols in their list of ten questions to answer

[1] Association of Threat Assessment Professionals, *Risk Assessment Guideline Elements for Violence: Considerations for Assessing the Risk of Future Violent Behavior* (Sacramento, CA: Author, 2006), https://c.ymcdn.com/sites/www.atapworldwide.org/resource/resmgr/imported/documents/RAGE-V.pdf.

[2] Association of Threat Assessment Professionals, 5.

[3] Frederick S. Calhoun and Stephen W. Weston, *Concepts and Case Studies in Threat Management* (Boca Raton: CRC Press, 2013), 11, http://www.crcnetbase.com.libproxy.nps.edu/isbn/9781439892183.

[4] Robert A. Fein and Bryan Vossekuil, *Protective Intelligence and Threat Assessment Investigations: A Guide for State and Local Law Enforcement Officials* (Washington, DC: U.S. Department of Justice, 1998), 24.

during assessments. Answering the questions requires tasks such as surveillance on the threatener and interviews with neighbors and friends.[5] Steps like these obviously require waiting until the identity of the threatener is known. Working on identifying suspects so an investigation can begin does not stop new threats and reposts of older threats, possibly breaking down a threatener's inhibitions for violent speech, but not for true violence.[6] Thus, waiting to start risk assessments is no longer acceptable. Providing security while identifying a suspect requires the use of limited resources. The ability of Twitter and Facebook to communicate to a large group of people quickly, possibly gaining a community of support, means law enforcement needs a new way to begin assessing the risk of violence sooner.

Analyzing the language of Twitter and Facebook posts could provide an earlier starting point. Patton et al. illustrate how language from Twitter messages can be coded for different types of aggression—direct, indirect, proactive, and reactive.[7] The coding method is not used to assess risk. Coding Twitter messages also occurs in the private sector. Automated analysis of social-media language can identify risks to a company's reputation through a coding process.[8] By coding text, analysts can divide tweets into negative or positive feelings about the company.[9] Progressing even further, confidence scoring provides a ranking based on how probable an event is to happen.[10] Using this tool could also help determine which threats have a higher probability of happening. Lastly,

[5] Randy Borum et al., "Threat Assessment: Defining an Approach to Assessing Risk for Targeted Violence," *Behavioral Sciences & the Law* 17, no. 3 (September 1999): 331–335, doi: 10.1002/(SICI)1099-0798(199907/09)17:33.0.CO;2-G.

[6] "Online Disinhibition Effect (Suler)," *Learning Theories*, December 15, 2015, https://www.learning-theories.com/online-disinhibition-effect-suler.html.

[7] Desmond U. Patton et al., "Gang Violence on the Digital Street: Case Study of a South Side Chicago Gang Member's Twitter Communication," *New Media & Society* (January 2016): 7, doi: 1461444815625949.

[8] Paul Alpar and Daniel Ohliger, "Creation of Risk Profiles of Business Customers from Social Media," *Banking and Information Technology* 16, no. 1 (March 2015): 26, http://web.a.ebscohost.com.libproxy.nps.edu/ehost/pdfviewer/pdfviewer?sid=a77475c8-66bb-44f6-9007-565450b5762d%40sessionmgr4007&vid=1&hid=4214.

[9] Alpar and Ohliger, 26.

[10] Bill Murdock, "How to select a threshold for Acting Using Confidence Scores," IBM Watson, June 23, 2016,https://developer.ibm.com/watson/blog/2016/06/23/how-to-select-a-threshold-for-acting-using-confidence-scores/.

tools such as public record databases search engines and social media analysis softwares exist which could analyze information provided in threatening posts, feed information to the confidence scoring tool and possibly increase the accuracy of confidence.

In sum, risk assessments begin at the point a suspect is identified. The evolution of social networking sites allows individuals to communicate threats anonymously and in high volume. This ability means investigators cannot wait to identify a suspect to begin a risk assessment. There are software tools already available that can help start assessments earlier. By creating a new software platform that combines a social media monitoring tool, a language sentiment tool, a criminal history database and a confidence scoring tool, law enforcement identifies violent people before they injure or kill their victims. Although more development is needed, the case stories in this thesis shows the proposed software correctly identifying people that post on social media and then act out violently.

C. LITERATURE REVIEW

Researching the topic of threat assessments led to articles in professional journals, guides by the Department of Justice, as well as peer-reviewed publications. Much of the material is decades-old; some publications date back to the late ‘90s. As a result, the publications do not discuss threats over Twitter and Facebook. However, publications by the business, communication, and computer-science fields do discuss risk assessments in the context of Twitter and Facebook. Expanding the search to fields outside the criminal justice profession necessitates breaking this literature review into three areas: threat assessments, social-media communication, and risk assessment methods. The first section includes scholarly publications from the 90s on threat assessment models. These models come from the field of psychology and fit into clinical assessment practices. The psychology of social-media audiences sets up the second section for academics to discuss how Twitter and Facebook make it easier for suspects to threaten others. The third section explores assessments of written terrorist and suicide threats as well as risk assessments in business. Research done on these subjects may help update the standards law enforcement currently uses.

1. Threat Assessments in Law Enforcement

Scholars write about threat assessments for use in a wide variety of threat environments. Cornwell et al. provide guidelines for assessing threats in the education field, Neben applies threat assessment models to lone-wolf terrorists, and Patton et al. analyze threats among gang members.[11] This section narrows the scope to threats against law enforcement. Narrowing the scope allows the discussion of threat assessment guides used by law enforcement as well as the psychology behind threats.

While working for the Secret Service, Fein and Vossekuil created a threat-assessment guide for law enforcement.[12] The U.S. Department of Justice published Fein and Vossekuil's guide as a reference for law enforcement officers tasked with investigating threats.[13] This exemplar provides descriptions of assassin behavior, elements of a threat assessment program, and guidance for conducting threat assessments.[14] In 1999, these authors joined Borum and Berglund to produce an academic work for the behavioral science field, outlining questions to answer during a threat assessment.[15] The ten questions developed in their article form a behavioral approach to assessments, which has earned the scholars frequent acknowledgments by academics and practitioners alike.[16] The questions help explain behavior—the subject's interest in violence, his communication of intentions for violence to friends and family, and his engagement in actions considered precursors to violence such as stalking.

Calhoun and Weston are two other authors frequently referenced in threat-assessment literature.[17] Their work, influenced by experience helping the U.S. Marshal

[11] Rachel V. Neben, "Effectiveness of Threat Assessment Models for Lone Terrorists," *Small Wars Journal* 13, no. 7 (August 2015): http://smallwarsjournal.com/jrnl/art/effectiveness-of-threat-assessment-models-for-lone-terror; and Patton et al., "Gang Violence on the Digital Street."

[12] Fein and Vossekuil, *Protective Intelligence*.

[13] Robert A. Fein, Gwen A. Holden, and Bryan Vossekuil, *Threat Assessment: An Approach to Prevent Targeted Violence* (Washington, DC: U.S. Department of Justice, 1995), https://www.hitacllc.com/HITAC_Resources/ThreatAssessmentApproachtoTargetedViolence.pdf.

[14] Fein, Holden and Vossekuil, 3–4.

[15] Borum et al., "Threat Assessment," 331–335,

[16] Borum et al., 331–335.

[17] Frederick S. Calhoun and Stephen W. Weston, *Concepts and Case Studies in Threat Management* (Boca Raton: CRC Press, 2013), 1, http://www.crcnetbase.com.libproxy.nps.edu/isbn/9781439892183.

Service and the California Highway Patrol investigate threats to government officials, describes a pathway to intended violence, which includes the stages of grievance, ideation, research and planning, preparation, breach, and attack.[18] Calhoun and Weston introduce the theory of "howlers and hunters as they discuss the movement from grievance to attack."[19] Howlers are satisfied with just communicating threats. Most of the time, they do not leave the ideation stage on the pathway; the thought of violence is enough to satisfy them.[20] Hunters quickly move down the path. There is no threat communication, and suspects reach the attack stage.[21] This theory helps explain why some people threaten and others attack without warning. The theory does not rule out threateners as attackers. Although Calhoun and Weston updated their book in 2013, it does not discuss threats via Facebook and Twitter.[22] This lapse by prominent authors shows the need to update threat assessment research.

2. Social Media Communication

Social-media literature dates back to when email and blogs were the main form of communication over the internet. In a 2016 article, Carpenter and Lertpratchya discuss general communication over current social-media platforms such as Twitter and Facebook.[23] They create a model illustrating social media as a "customer service provider, mobilizer, information disseminator, researcher, and community builder."[24] Carpenter and Lertpratchya expand on the community builder role, describing how social media is malleable to fit individual needs to connect them to their online community.[25] The community-builder role shows up in other research as academics discuss the effects of social media on people's lives.

[18] Calhoun and Weston, *Concepts and Case Studies*, 10.

[19] Calhoun and Weston, 11.

[20] Calhoun and Weston, 10–11.

[21] Calhoun and Weston , 11–12.

[22] Calhoun and Weston.

[23] Serena Carpenter and Alisa P. Lertpratchya, "Social Media Communicator Roles: A Scale," *Social Media + Society* 2, no. 1 (January–March 2016): doi:10.1177/2056305116632778.

[24] Carpenter and Lertpratchya, 1.

[25] Carpenter and Lertpratchya, 7.

Dijck and Poell describe how social media platforms are pervasive in people's lives.[26] Health care, education, and even civil protests all use social media.[27] Dijck and Poell along with Shephard et al. discuss the growth and sustainability of political action spawned over social media while discussing topics such as "cloud protesting" and "hate speech."[28] The barrier to protesting and hate speech is lowered on social media because it provides a level of anonymity that lowers inhibitions. This effect ends up a subject of a roundtable discussion on the topic of spreading hateful speech throughout a community over social media.[29] Masur and Scharkow write that social media "connects people, applications and business."[30] This ability to bridge people leads to a discussion of the level of relationships individuals believe they have on social media and how it drives their perception on privacy of their posts.[31] Twitter and Facebook make it difficult to keep different social groups from overlapping and to control content between them.[32]

For another roundtable, academics in the fields of communications and media discuss the pervasiveness of social media.[33] Shepherd et al. discuss how hate has evolved online.[34] They emphasize how easy it is to post hate speech and threats online versus in person.[35] The participants also suggest how difficult it is to determine what language ends in violence.[36] Part of this difficulty is in the varying interpretations of language in a

[26] José van Dijck and Thomas Poell, "Social Media and the Transformation of Public Space," *Social Media + Society* 1, no. 2 (2015): 1, doi: 10.1177/2056305115622482.

[27] van Dijck and Poell, 1.

[28] van Dijck and Poell, "Social Media and the Transformation of Public Space," 3; and Tamara Shepherd et al., "Histories of Hating," *Social Media + Society* 1, no. 2 (July–December 2015): 1, doi:10.1177/2056305115603997.

[29] Shepherd et al., "Histories of Hating," 1–10.

[30] Philipp K. Masur and Michael Scharkow. "Disclosure Management on Social Network Sites: Individual Privacy Perceptions and User-Directed Privacy Strategies." *Social Media+ Society* 2, no. 1 (2016): 1, doi: 10.1177/2056305116634368.

[31] Masur and Scharkow, 3.

[32] Yumi Jung and Emilee Rader, "The Imagined Audience and Privacy Concern on Facebook: Differences Between Producers and Consumers." *Social Media + Society* 2, no. 2 (2016): 1, doi:10.1177/2056305116644615.

[33] Shepherd et al., "Histories of Hating."

[34] Shepherd et al., 1.

[35] Shepherd et al., 2.

[36] Shepherd et al., 5.

post.[37] While Sheppard et al. discuss hate and threatening speech on social media, their discussion does not provide answers for assessing the threats.[38]

Patton et al. do propose a method for assessing threats communicated over Twitter.[39] This limited study concentrates on threatening communication among gang members in Chicago, leading to the death of a gang member named Tyquan Assassin.[40] Tweets from a two-week period surrounding her death were coded into categories of violence such as direct threats of violence and indirect threats of violence.[41] Patton et al. show that Twitter messages convey geographical information for locations of violence as well as the mechanism for that violence.[42]

3. Risk Assessment Methods

Research on threats communicated over social media fails to address threat assessments over Twitter and Facebook in the law-enforcement community. In an attempt to gather information to address the deficiency, literature on evaluating risk in the language of suicide notes, the written words of terrorist communications and social media language directed at businesses is explored.

Handleman and Lester use linguistic inquiry and word-count analysis programs in their study of suicide notes.[43] Their methods looked at documents word-for-word and analyzed the text against 70 different aspects.[44] They find the words people choose can show distress as well as feelings of inclusion or exclusion from their social groups.[45] The ability to look at written language for these signs may identify people needing help prior

[37] Shepherd et al., 4–5.

[38] Shepherd et al., 5.

[39] Desmond U. Patton et al., "Gang Violence on the Digital Street," 6.

[40] Patton et al.

[41] Patton et al., 7.

[42] Patton et al., 12.

[43] Lori D. Handleman and David Lester, "The Content of Suicide Notes from Attempters and Completers," *Crisis* 28, no. 2 (2007): 102, doi:10.1027/0227-5910.28.2.102.

[44] Handleman and Lester, 102.

[45] Handleman and Lester, 104.

to committing suicide.[46] The ability to apply this method to Twitter posts still needs evaluation.

Literature on terrorist communication uses a different method to evaluate threats. Value references, words choice showing how someone values something, in written communication are the center of communication research in this field. Smith et al. investigate how word choice demonstrates terrorist groups feelings about adversaries.[47] Value references within written communications of terrorists reveal which terrorist groups are closely associated as well as which groups are outcasts.[48] This method of word analysis requires coding and rating of the document.[49]

The business field also uses coding language. Literature in this field focuses not on violence but on the effects of reputation. Sipior et al. postulate that businesses focus on social media's positive effects.[50] A number of articles spanning the disciplines from business to communications discuss the use of social media to improve a company's reputation.[51] However, the academics do not consider the risk of social media to a company's reputation.[52] Stepashkin and Khusnolarov present a way to begin analyzing risk to reputation through an automated process.[53] Their process involves establishing a special language based on semantic categories and lexicons.[54] Once the language is

[46] Handleman and Lester, 104.

[47] Allison Smith, "From Words to Action: Exploring the Relationship between a Group's Value References and Its Likelihood of Engaging in Terrorism," *Studies in Conflict and Terrorism*, 27, no. 5 (2004): 409, doi:10.1080/10576100490483679.

[48] Smith., 412– 413.

[49] Smith, 420.

[50] Janice C. Sipior, Burke T. Ward, and Linda Volonino. "Benefits and Risks of Social Business: Are Companies Considering E-Discovery?." *Information Systems Management* 31, no. 4 (Fall 2014): 328, doi:10.1080/10580530.2014.958031.

[51] Mark Brinkley, "Social Media Risk." *Internal Auditor* 71, no. 2, (April 2014): 68– 69, http://web.a.ebscohost.com.libproxy.nps.edu/ehost/detail/detail?sid=2b2cf9a3-e143-4905-8609-a6c06b9b3d8d%40sessionmgr4007&vid=0&hid=4209&bdata=JnNpdGU9ZWhvc3QtbGl2ZSZzY29wZT1zaXRl#db=bth&AN=100244213; and Sipior et al., "Benefits and Risks of Social Business."

[52] Sipior et al., 331.

[53] M.V. Stepashkin and F.F. Khusnolarov, "Risk Analysis for Reputation Based on Assessments and Ranking of Information Events and Specific Data from Open Sources of Information." *Problems of Economic Transition* 57, no. 12, (2015): 8, doi:10.1080/10611991.2015.1161443.

[54] Stepashkin and Khusnolarov, 11.

established, the researchers create templates for groups of communications.[55] These templates help divide the language into positive, negative, and neutral categories.[56] Finally, people trained on their coding system assess the fact a second time, providing confirmation.[57] Stepashkin and Khusnolarov do not discuss the accuracy of the mechanism.[58]

4. Conclusion

The use of social media as a venue to communicate threats is a relatively new field of study. Therefore, research on assessing threats delivered over social-media platforms, such as Twitter and Facebook, is still catching up. While the literature covers techniques that are useful in assessing and mitigating threats, specific applications to communications over social media is not adequately covered. The case study by Patton et al. on gang communications over Twitter and Facebook provides a starting point and a methodology of inductive textual analysis that is useful for coding violence.[59]

D. RESEARCH DESIGN

Twitter and Facebook allow the communication of threats quickly, anonymously and in high volume. This presents a situation where law enforcement has a large number of threats to investigate. Current methods for assessing the risk from a threat start at the point an individual are identified. This means investigators sort threats by ease of identification instead of risk for violence. Prioritizing threat investigations by intuition alone could lead to not investigating a threat, resulting in a victim being injured or killed. Currently, there is not a better way to prioritize the threats. If a suspect is not known, the language of the post is the only avenue available to determine risk of violence. Researching whether a model applied against the language of Twitter and Facebook posts can determine the confidence of a threat happening may allow investigators to correctly

[55] Stepashkin and Khusnolarov, 11.

[56] Stepashkin and Khusnolarov, 11.

[57] Stepashkin and Khusnolarov, 11.

[58] Stepashkin and Khusnolarov, 15.

[59] Patton et al., “Gang Violence on the Digital Street,” 6.

prioritize investigations. This could help prevent injury or death to victims of violent threats.

This thesis aims to propose the model that can start assessing the risk of threats communicated over social media before a suspect is identified. The proposed model being assessed combines software platforms already used by law enforcement and assesses how they can combine with software known as confidence scoring to automate analysis and provide a confidence of occurrence for threats. Software requirement specifications used in the proposal follow IEEE 830 -1998 format. IEEE 830–1998 is the recommended format for communicating software requirements. The paper also compares current investigative practice against the proposed software in threat investigation and demonstrates how this model can be implemented and enhances threat investigations. This comparison is enhanced through exploration of four case stories. Each story walks the reader through how investigations manually make case decisions. In a later chapter, the step within the proposed software evaluates the same case stories and illustrates how the software aids investigations.

The research design sets up an intuitive progression for this paper. Chapter II explains how current investigators work social media threat cases. After the explanation of the investigation method, four case studies show how detectives apply the steps. Chapter III then describes how the proposed software works threat investigations. This is accomplished through completed a software requirement specification document that becomes chapter III. Once the explanation is complete, the same previous case studies show how the way social media threats investigations improve. While new software to help improve investigations is encouraging, the development does not come without areas of consideration. Chapter IV explores implementation and design obstacles and wraps up with the thesis conclusion. Upon this conclusion, this thesis hopes to demonstrate an area for future research in how coding language can change to allow each to work together in a user friendly product.

THIS PAGE INTENTIONALLY LEFT BLANK

II. CURRENT INVESTIGATIVE METHOD

Threats over social media create a situation whereby multiple threats can be communicated anonymously. While assessing these threats starts at discovery, and includes deciding on protective measures for the victim, efforts of investigators center on identifying the suspect. The steps to identifying the poster take time and force investigators to use their best judgement for what order they investigate the threats. The order investigators chose to investigate cases can end with more serious ones being overlooked. The following sections illustrate how investigators verify a threat and search for suspect identification in order to complete a risk assessment.

A. THE MANUAL INVESTIGATIVE PROCESS

When a threat is identified on Twitter or Facebook, it is assigned an investigator. The investigator's first step is to verify that the message is a threat. He then begins the process of identifying a suspect. If the threat is posted with the public privacy setting, the investigator can go directly to the online post. The investigator can access the user name for the account which is sometimes the user's real name and sometimes a fake name. Accounts with fake user information or whose security is set to private require the investigator to appear before a judge for a court order. If the judge issues an order, it mandates that Twitter or Facebook preserve the posts and handover account information to the law enforcement investigator. This process takes time to write the court affidavits, secure a judge's signature, submit the order to Twitter or Facebook, and wait hours or days for the social media provider to send the requested information.

Twitter or Facebook may only provide fake information. Additional court orders or search warrants may compel Twitter and Facebook to provide internet protocol (IP) address information for the threatening post. This second court orders provide a name for the internet provider. A third court order or search warrant is then submitted to that internet provider requiring them to provide the location of the IP address. Database research by the investigator may reveal a suspect living at the address. Surveillance of the address may assist in suspect identification. A search of the address after obtaining a search warrant or interviewing the occupants of the address may lead to the identity of suspects. If a suspect is identified, a risk assessment can begin.

Even if a suspect is identified quickly, during times of high threat volume, investigators must decide which threats to put limited manpower toward. There is no standardized guidance for how to make these decisions. The four case studies that follow analyze two cases the Raleigh Police Department investigated during an officer involved shooting incident in February 2016 and two cases that happened over a year later and each from different jurisdictions.

B. CASE STORIES

In this section, three posts are examined through the lens of three criteria: social media language, the identity and criminal background of the poster, and the statutory requirements to deem the post a threat. Each case includes the language for the posted conversational thread and discusses how the language, the criminal background of the poster and criminal statutes are used to evaluate the post for a threat. The authors of the first two case studies posted their comments in open source; however, due to the investigations of their cases not ending in criminal charges, their names are changed to protect their privacy. These first two studies also originate from among 22 cases that investigators worked during the February 2016 incident. The last two studies stand alone and are from separate jurisdictions with one originating in Washington, D.C., and the other from New York City. The language for case story three and four come from media sources.

Expletives remain written as they were in these news stories. The original spelling and punctuation of each social media was unchanged from the original.

1. Facebook Post against Officer Twiddy

In February 2016, a Raleigh Police officer shot a suspect during a struggle. Shortly after the incident, people from the community started posting on Facebook and Twitter. A lot of these posts talked negatively about the police. Other posts communicated threats toward law enforcement. This case study is taken from a Facebook post made after the shooting. The person posting was identified and referred to here as B.K. He posts the following comment:

> Im sorry I don't normally think about this but peaceful protesting? For what? For them to continue to take us African Americans as a joke. I may be wrong but the way I'm feeling it's time to shoot some of they kind down too. We always talk about other cities going through what we going through && what we would do. I'm going to support my city 100% RIP Lockman.[60]

This post is followed by a comment from another subject wanting peaceful protests in his neighborhood, so neighbors' houses and kids do not get shot. B.K. replies to this individual with

> Ain't nobody thinking about shooting up nobodies house or kids….I said shoot some of their kind down not go shoot up houses and kill kids.[61]

B.K. then posts on his Facebook page a picture of two t-shirts, one reading "Fuck a cop named Twiddy" and the other "100 to 500 RIP Lock AKA."[62]

a. Social Media Language

The language in this post expresses direct violence against law enforcement officers, and the direct action was to shoot officers. When faced with a poster asking for peaceful protest, B.K. continues with the threat of violence. B.K. then posts pictures of t-shirts naming a specific officer and encouraging people to load an "AKA," an assault rifle. Moreover, this post appeared during a period when the community was already protesting the shooting of a black male by the police. B.K. did fall into the *howler* definition in that he posted in anger. Most howlers do not turn to violence, but since his claims were specific and communicated more than once, B.K. warranted a closer look.

[60] B. K. Facebook page, accessed March 1, 2016, https://facebook.com/prfile.php?id=100010733497326&fref=ts.

[61] B. K. Facebook page, https://facebook.com/prfile.php?id=100010733497326&fref=ts.

[62] B. K. Facebook page, https://facebook.com/prfile.php?id=100010733497326&fref=ts.

b. Criminal Background

Once the poster is identified, investigator completes a criminal background check. This can help show any propensity toward violence or threatening behavior. B.K.'s criminal record revealed nothing indicating a history of communicating threats, harassment, or violence. Since B.K. lives in the community, a local background check is made and shows no involuntary commitment (IVC) for mental evaluation. Since B.K. does not have any criminal history, he may be blowing off steam as a howler.

c. Statutory Requirement

The third criterion determines whether the post fits the statutory language of a threat. B.K. willfully communicated a threat of violence, identified a distinct group of people through his threat, and implied a specific officer. Investigators could criminally charge on this post under the third criterion alone. Factoring in the heightened emotions after the incident, B.K.'s lack of criminal history, and his threatening language focusing on police as a whole, the prosecution of the charge would be difficult. B.K.'s threats did not continue beyond the third post, after which he reverted to pre-threat language. Because his language reverted back to that of pre-incident posts, analysts felt that B.K. was a howler. Detectives did not charge, the poster did not commit a violent act.

2. Facebook Post against the Raleigh Police Department

The second case study also looks at a post from the February 2016 shooting incident and explores a series of posts made by T.W. spanning March 3 to March 6, 2016. T.W.'s identity was known. On Facebook, T.W. writes,

> […] if RPD [Raleigh Police Department] lies one more time, I'm setting the whole shit on fire tonight." Several comments to this post tried to dissuade T.W. from using fire. T.W. replied with "… assembling today to show force. Meet at 5pm. […][63]

T.W.'s next post is "I can't continue to be docile. .. I am ready to ride or die (bomb, gun, bomb [emoji]) I WILL GO ALONE IF NEED BE."[64] Three people post comments telling

[63] T.W. Facebook page, accessed March 2, 2016, https://facebook.com/mr.nething.

[64] T.W. Facebook page, https://facebook.com/mr.nething.

him not to resort to violence. T.W. writes, “They’re leaving us no choice,” which received the comment from another poster “we always have choices.”[65] T.W. responds, “You have to be willing to die for what you believe in.”[66]

a. Social Media Language

T.W.’s language describes a direct action when he writes about “setting the whole shit on fire,” but he prefaces that action with certain criteria that must be met before the action happens. The rest of T.W.’s language is nonspecific, implied violence. Neither law enforcement nor any one individual is named in his threat. It is not clear what he is threatening, but T.W. is posting language to be heard. He seeks support, but as seen in the comments to his posts, no one latches onto his idea. The conversation lasts for three days, longer than any other thread, before T.W. returns to pre-event language.

b. Criminal Background

T.W.’s criminal record shows nothing in line with communicating threats, harassment, or violence. There has also been no contact with mental health professionals recorded by local law enforcement. The language is concerning, but nothing in the poster’s background shows threatening behavior or violence.

c. Statutory Requirement

There are not enough specifics in his language to meet the elements of communicating a threat. The non-specific threat tied to a criterion lowers the fear of the violence actually happening. Nonspecific acts also affect the perception that the threatener is really able to carry out any action. No further action is taken by police.

The first two studies contain language consistent with others authored during the incident. Detectives discarded re-posts which allowed them to narrow the cases to 22 rather quickly. All 22 cases contained the type of language that is illustrated in the two case studies. By utilizing the three criteria—language, criminal history, and statutory

[65] T.W. Facebook page, https://facebook.com/mr.nething.

[66] T.W. Facebook page, https://facebook.com/mr.nething.

requirements—Raleigh Intelligence Center (RIC) detectives were able to triage posts to determine how far to investigate them. Since community tensions were high against the police, these criteria allow the detectives to conclude that community contact through interviewing or taking other inhibiting measures was not needed. None of the 22 online threats were acted on by the posters, and each poster returned to pre-incident language.

3. Facebook Post to the Republican Party

James T. Hodgkinson shot members of the Republican Party as they practiced baseball on June 14, 2017. He came to Washington, DC, from Illinois shortly after the presidential election and lived out of his van. The majority of his Facebook posts focused on the results of that election, primarily his dislike of the Republican Party. This case study examines some of those Facebook posts and starts with a post from March 22, 2017, that reads, "Trump is a Traitor. Trump Has Destroyed Our Democracy. It's Time to Destroy Trump & Co."[67] Other posts include one that reads

> I Want to Say Mr. President, for being an a** hole you are Truly the Biggest A** Hole We Have Ever Had in the Oval Office,

posted on June 12, 2017, and another that reads

> Republican B**ch [a reference to Republican Karin Handel on her comment that she doesn't support a living wage] Wants People to Work for Slave Wages, when a Livable Wage is the Only Way to Go! Vote Blue, It's Right for You!

posted on June 8, 2017.[68] Hodgkinson's identity and posts came from media sources reporting on the shooting.

a. Social Media Language

The language of these posts does not convey direct threats. The post on March 22, 2017, uses the word "destroy." This term can have multiple meanings. It can mean something like bringing down the party by voting them out-of-office or something as

[67] "Saved From https://www.facebook.com/jthodgkinson," Archive.is Web Capture, accessed September 15, 2017, http://archive.is/QH4A8#selection-2751.0-2751.87.

[68] "Saved From https://www.facebook.com/james.hodgkinson.568," Archive.is Web Capture, accessed September 15, 2017, http://archive.is/OncTJ#selection-5909.0-5909.128.

extreme as killing them one by one. This broad of meaning makes it difficult to say he is threatening physical harm to Trump and Company. The phrase may provide investigators reasonable suspicion to take a deeper look at the poster but without more posts helping to clarify the meaning of the "destroy Trump and Company," the post is not a threat of violence. In this case, Hodgkinson continues to post about the Republican Party. He uses offensive words such as asshole and bitch to describe the president and his displeasure in the platform. None of his posts comment further on any action he is planning to take that helps define his meaning of the word "Destroy." None of his posts meet the definition of a threat.

b. Criminal History

The Facebook poster is readily identified as James T Hodgkinson. The word "Destroy" in the post provides reasonable suspicion to look at Hodgkinson's criminal history. His criminal history shows one incident resulting in two assault charges and one aggravated discharge of a firearms charge. This information is useful, but since the wording of the posts did not rise to a direct threat, this one incident would probably not raise much concern.

c. Statutory Requirement

The language of the posts do not meet statutory requirements of threats. None of the posts say that Hodgkinson is going to hurt anyone. The post saying it is time to destroy Trump and Company is not detailed enough and does not meet the other statutory requirement of fear that the act can really happen. Even with a criminal history of assaults, without the poster meeting the language requirements, a district attorney would not have pressed charges. This is the primary reason why criminal history alone does not cause much concern for investigators.

4. Facebook Post to New York Police Officers

On July 5, 2017, John Bonds shot and killed a New York City police officer while she sat in a command vehicle. The shooting appears unprovoked as video of the crime shows Bond walking from a store and along the wall toward the command vehicle.

Shortly after the shooting, a Facebook Live video from September 2016 shows Bond talking harshly about the police. The media posted this video during the reporting on the incident. As with the other three case studies, this study looks at Bonds' video through its language, the poster's criminal history and the statutory requirements to call it a threat.

The *Wall Street Journal* reports that Bonds made the video in September 2016, and it contained the following language:

> I'm not playing, Mister Officer. I don't care about a hundred police watching this s—shit. You see this face. You see this face or anything, leave it alone. Trust and believe, [...]
>
> I'm not hesitating. It ain't happening, […..] I wasn't a b—- bitch in jail, and I'm not going to be a b—- bitch in the streets.[69]

Other comments from this video are documented by the *Daily News* as

> Y'all n-----s so reluctant to want to say something to the police, man,
>
> Man, police is f----ts, and this ain't no gimmick. F----ts. [...] N-----s ain't taking it no more, Mr. Officer. I'm here to tell you, man. ... just keep your a-- away from mine.[70]

a. Social Media Language

The language in this video does not threaten violence. Bond's emotionally narrates as if talking directly to a police officer. He angrily tells them that they should not mess with him; that he wants to be left alone. Bond's word "I'm not hesitating" seems to imply he will act if the police mess with him but he does not qualify that comment with a type of action. His rant ends by repeating his plea to the police to stay away from him. This still does not rise to a direct threat against anyone. This video appeared among

[69] Kristen Phillips, Mark Berman, and Wesley Lowery, "I'm Not Playing, Mr. Officer': Gunman Appears to Complain About Police Mistreatment in Video Months Before Shooting NYPD Officer," *Washington Post,* July 5, 2017, https://www.washingtonpost.com/news/post-nation/wp/2017/07/05/assassinated-nypd-officer-shot-and-killed-while-sitting-in-a-police-vehicle-officials-say/?utm_term=.a1962cdcfc15.

; and Graham Rayman, and Larry Mcshane, "NYPD Cop Killer Alexander Bonds Posted Anti-Police Facebook Rant," *Daily News*, July 5, 2017, http://www.nydailynews.com/new-york/nyc-crime/suspected-nypd-shooter-assaulted-officer-brass-knuckles-article-1.3302356.

Bonds' other posts, which are mostly inspirational quotes.[71] Officer routinely get told that people want to be left alone so these statement are not uncommon.

b. Criminal History

Bonds has a lengthy criminal history. He was paroled in 2013 after a prison sentence for robbery.[72] His record also includes charges of assaulting an officer with brass knuckles.[73] News reports do not indicate whether Bonds has any IVC, but ABC News reports his girlfriend as saying he was not taking his psychiatric medicine, implying that he had mental illness.[74] The criminal record and mental illness is concerning, but the type of ramblings are not uncommon among people that have had multiple arrests by the police, particularly when they have mental illness. A person making these comments would not cause concern past putting out an officer safety bulletin for officers to approach him with caution.

c. Statutory Requirements

The language of Bond's posts do not indicate violence toward any specific officer or officers. He does not talk about any specific violence at all. He does say that if he is not left alone, that he will do something but that something is not identified. Without more specific details, the language does not meet statutory requirements to charge Bonds with a crime.

C. CONCLUSIONS

The cases in this chapter came from three incidents involving four unacquainted individuals. Two of the individuals attacked people after their concerning posts. Hodgkinson wrote letters to political representatives and posted constantly the days leading up to him shooting a senator. These posts are very anti republican but do not

[71] Colleen Long and Jennifer Peltz, "Officer's Killer had Ranted About Police Killing and Abusing," *ABC News,* July 5, 2017, http://abcnews.go.com/US/wireStory/female-police-officer-critical-shooting-bronx-48444782.

[72] Long and Peltz.

[73] Long and Peltz.

[74] Long and Peltz.

threaten violence. Bond posts an emotional video but no violence follows immediately. His Facebook page contains inspirational quotes until he shoots a police officer ten months after the video for police to leave him alone. With both Hodgkinson and Bond, it is unclear whether law enforcement knew about their posts before the attacks. Raleigh police learned of the posts for the first two case studies shortly after they appeared on Facebook. None of the four case studies contains language that meets statutory requirements of a threat. Interestingly, the examples do show that violence can follow posts that do not meet the requirement of a threat.

Police intervention in Facebook and Twitter post cases is difficult. Investigators are first hindered by statutory requirements needed for them to use the legal system to make criminal charges. The language of the post must directly identify who is targeted and make those people feel like the threat is really going to happen. As the case studies illustrate, the language is usually not that specific. The second option the police have is to go talk to the person authoring the post. Sometime just talking to the person can be successful in deterring violence, or it can further infuriate the person. Knowing this places investigators in a position where they must make a subjective decision. This causes situations where investigators may decide not to do anything. They cannot charge because the language of the threat does not meet the statutory requirements and they feel that talking to the subject may make him angrier. This causes a situation where cases that need intervention may not receive it.

III. AUTOMATING TWITTER AND FACEBOOK THREAT INVESTIGATIONS

The previous chapter showed how current social media threat investigations work. This process is slow, resource intensive, and subjective. Using technology could speed up investigations, lower resource needs, and improve objectivity. In an attempt to move investigative practices toward using technology, this chapter takes on a different format. Proposing new software requires utilizing a technical format to communicate the needed software functions to a developer. These technical instructions used to be long detailed documents, but in agile business communities, they are very concise.[75] In this chapter, a software requirement specification (SRS) document details a tool to assist investigators in locating a threat, analyzing the threat language, running a poster's criminal background, and ultimately deciding which cases investigators should investigate. The sections consisting of an introduction to the software, a description of the software, and case stories on the use of the software show the functional requirements. Each of these sections complies with a suggested standard designed by the Institute of Electrical and Electronics Engineers (IEEE). The opening letter of IEEE 830–1998 provides a good explanation of the recommended standards and their use:

> IEEE Standards documents are developed within the IEEE Societies and the Standards Coordinating Committees of the IEEE Standards Association (IEEE-SA) Standards Board. The IEEE develops its standards through a consensus development process, approved by the American National Standards Institute, which brings together volunteers representing varied viewpoints and interests to achieve the final product....Use of an IEEE Standard is wholly voluntary… Every IEEE Standard is subjected to review at least every five years for revision or reaffirmation. When a document is more than five years old and has not been reaffirmed, it is

[75] Jerry Cao, "A Practical Approach to Functional Specifications Documents," Studio, accessed September 15, 2017, https://www.uxpin.com/studio/blog/practical-approach-functional-specifications-documents/.

> reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art...[76]

The SRS in this document stays consistent with the intentions of the IEEE but does include some variation. Case stories, not specifically mentioned in the IEEE standard, conclude the SRS as a way of providing developers a way to visualize the complexity of the programing needed to develop this tool.

Including the case stories in the SRS gives developers and future users four scenarios that illustrate how each part of the system must work together in producing a threshold confidence score. Each of these case stories provide different characteristics that need considered to calibrate the software correctly. Case stories one and two exhibit language closer to direct threats than cases three and four. Case stories three and four, however, are the ones that end in violence. Both of the individuals did have crimes involving violence in their past while case studies one and two did not. The correlation between threat language and criminal history will need further study but could affect how the coding provides scores for these areas. The criminal history may need heavier weighting than the sentiment and wording of the language. The case studies demonstrate how a heavily weighted criminal history affects the confidence score and suggests that investigators look into the individual's posting further. If this software were to have helped locate and alert investigators to the posts that met threshold limits, intervention strategies may have changed the outcomes of the two cases that ended in violence.

Another area not illustrated in the case stories is the occurrence of false positives and false negatives. False positives related to threat investigation happen when the software identifies threats that meet the confidence threshold, but when assigned to a detective, the detective decides there is no threat. This is a checks and balance system for the software. Detectives take the information given by the software and apply their intuition and knowledge of their individual communities to make a final decision. The outcomes still feed back into the software and influence machine learning, which helps

[76] IEEE. *IEEE Std 830–1998 IEEE Recommended Practice for Software Requirements Specifications.* IEEE Computer Society, 1998.

continually calibrate the software outputs. The other occurrence is a false negative. A false negative happens when a post's confidence score does not trigger an alert and violence still happens. These scenarios provide information for machine learning, hopefully allowing better calibration for future events. These types of events also illustrate Calhoun and Watson's hunter category for people that carry out violence.[77] A hunter does not provide clues of an impending attack. The software not alerting on their posts provides further evidence of hunters' ability to stay below the radar of authorities.

While the outcomes of false positives and negatives can help the software learn and calibrate itself, a concern of the software over-calibrating to the point of infringing on free speech develops. Case story three shows some of these concerns. The language in Hodgkinson's posts is common for people posting their political disagreement. The software in this case takes into account the escalation, the negativity, the amount of posts, and the poster's arrest record before calculating a confidence score that triggers an alert. Software performing this step protects citizens from violations of their free speech. Law enforcement does not get to see any of these posts until the total score triggers the threshold alert. Setting the threshold level requires accuracy in determining the likelihood of violence. Reaching the threshold signifies that there is reasonable suspicion violence may happen. This is the same threshold law enforcement must have to investigate posts.

In these case stories, subjectivity still influenced the scoring. The scorer knew the outcomes of each story prior to the scoring. It is possible an investigator looking at these cases as they are playing out would score them differently. A larger data set of threats needs to be analyzed to align how each category scores. Machine learning takes the subjectivity away from people. The software learns from the outcomes of each threat it identifies. Adjustments to the algorithm improve the accuracy in identifying threats that end in violence. This ability helps take the subjectivity out of investigators' decisions. As shown through case story comparisons in Chapter II and the SRS in Chapter III, the automated method may allow law enforcement intervention before violence happens.

[77] Calhoun and Weston, *Concepts and Case Studies*, 11.

Software Requirement Specification
for a Social Media Threat Assessment Tool

1. INTRODUCTION

1.1 Purpose

The purpose of this document is to provide the details for creating software to assess social media threats. It explains how social language sentiment software, a public records search engine, and a confidence scoring tool interface to decide which social media posts warrant investigator intervention to prevent violence. This document details the software interfaces as well as the system's functions, and constraints to its operations. Case stories contained in this document help clarify the functional requirements of the software for both developers and end users. Part of this clarification contains limitations and obstacles that programmers must overcome.

1.2 Product Scope

The goal of the software is to help law enforcement investigators discover threats made over social media to start assessments and interventions prior to acts of violence. As a step to reach this goal, subjectivity needs to be removed from investigations. Machine learning and historical events function to objectively identify social media language that ends in violence. The software also protects citizens by identifying only the posts that meet a threshold for the possibility of violence. No other posts are reported to law enforcement.

This software will use confidence scoring to identify social media threats for law enforcement investigators. The software locates threats posted on social media applications through keyword searches and sentiment analysis tools. It then locates any poster identification information, compares it to criminal history records, and sends both the sentiment analysis and the criminal history information to a confidence-scoring tool. If the confidence-scoring tool results in a score that meets or exceeds a predetermined threshold, the software alerts the user to the threat. After the alert, the software generates a report that provides investigators information to further assess the risk for violence.

Each component in this software package already exists. The uniqueness of this new system is in the ability of the different components to work together. The other unique ability is after initial input from developers, the software operates independently of investigators, alerting only to threats that reach the predetermined threshold.

1.3 Definitions, Acronyms and Abbreviations

User	Someone who interacts with the software
SQL	Structured query language
API	Application program interface
RMS	Report writing management system
CAD	Computer aided dispatch
SRS	Software requirement specification

1.4 References

IEEE. IEEE Std 830–1998 IEEE Recommended Practice for Software Requirements Specifications. IEEE Computer Society, 1998.

1.5 Overview

The rest of this document includes three sections. Section 2 contains an overview of system functionality and interactions between other softwares contained in the system. This section also talks about the users and their interactions with the system. Finally, it explains the constraints and assumptions of the software. Section 3 provides the specification requirements in more detail and in terms that both a developer and an end user will understand. Section 4 presents a requirement analysis through the use of user stories. These stories help the developer visualize the system's functionality. These stories also explain to the law enforcement end user how the software determines which threats trigger the threshold alert.

2. GENERAL DESCRIPTION

In this section, hypothetical software is examined that assists investigators with locating a threat, analyzing the threat language, running a poster's criminal background, and ultimately deciding which cases investigators should investigate. To complete these tasks, this section explains how a social language sentiment software, a public records search engine, and a confidence scoring tool interface to decide which social media posts warrant intervention by investigators to prevent violence.

2.1 Product Perspective

The threat software is primarily self-contained software, operating with cloud technology. Once coded and calibrated by software and law enforcement teams, the end user requires limited ability to make adjustments to the algorithms. Users do interact through a web-based application. This means that clients do not need to invest in additional hardware. The software should be developed with the potential for end users to access software alerts over mobile devices such as smartphones. The software must either interface with or encode three existing software capabilities: social media analytics, language sentiment analysis, and confidence scoring. Connection with criminal history information requires an external interface to data held by law enforcement agencies or public record database companies. Figure 1 below illustrates the work-flow of the software

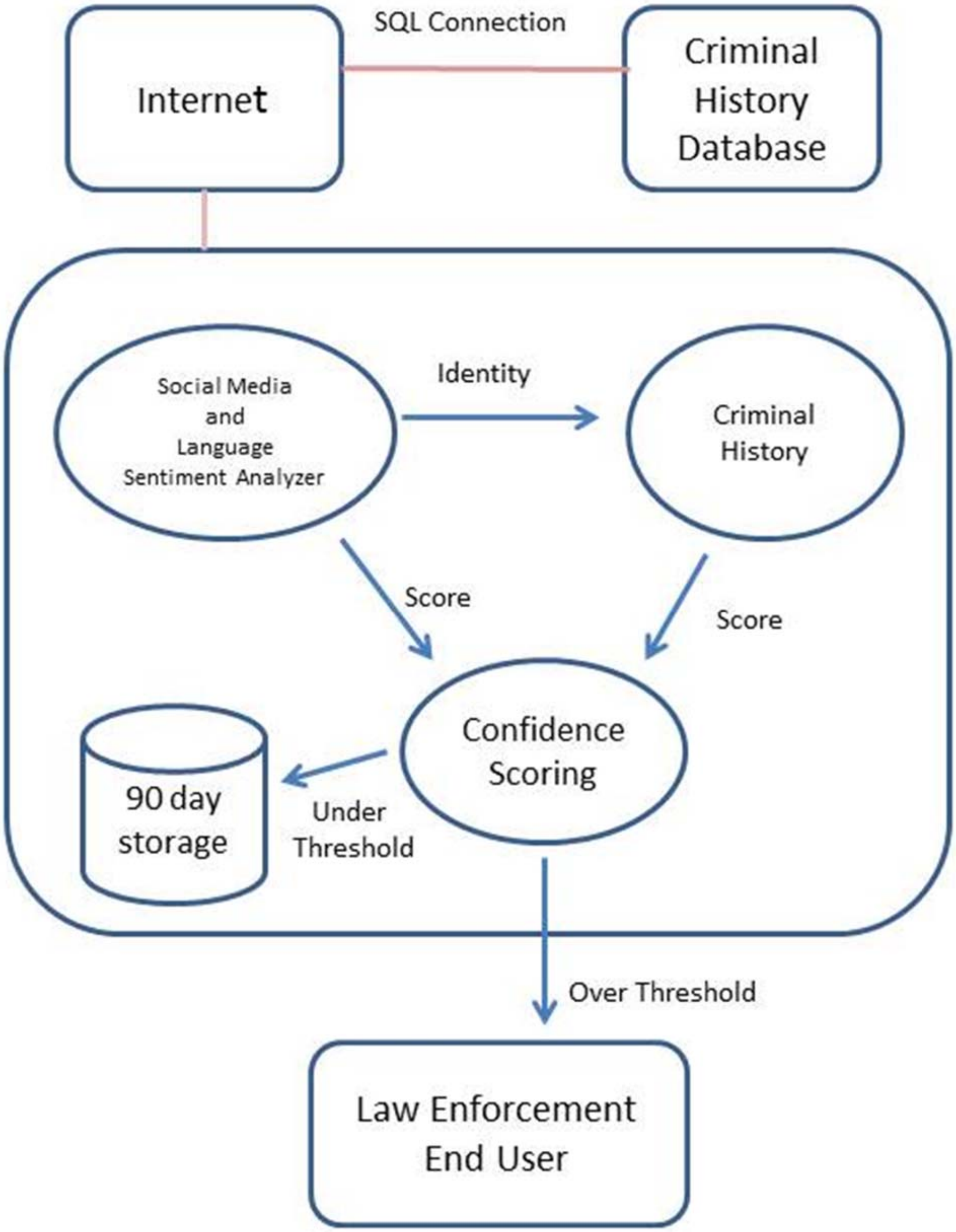

Figure 1. Block Diagram of Social Media Assessment Tool.

2.1.1 System Interfaces

The system is self-contained and accessible over a web browser. The system interfaces with an end user as well as other software and databases. These interfaces happen via a network-to-network connection through the internet as well as by connecting to databases through the use of SQL.

2.1.2. User Interfaces

There are two users who interact with the software. The development team designs and calibrates the system. They interact through the coding of the software and have the ability to input data through code to continually calibrate the system. Including the ability of machine-learning decreases the need for manual entry of calibrating data.

The second user is the law enforcement investigator. This category of users accesses the software through a web portal. They have the ability to access alerts and reports generated by the software. The users do not have access to other functions of the software to protect the objectivity.

2.1.3 Hardware Interfaces

The software is cloud-based, housed within a company's controlled servers. Connections to investigators' computers happen over internet connections from these servers to individual computers, each computer logging into the software remotely. Smartphone access allows alerts to reach investigators when away from their desks. This connectivity is also over a log-in through an app on the smartphone that remotely logs into the company-controlled software servers which has the software on it.

2.1.4. Software Interfaces

The system contains multiple applications already commercially available. Interfaces need developing that allows an existing social media analyzer to include a language sentiment algorithm. Another interface allows the data from the social media analyzer to pass to the confidence-scoring applications. Separately, an interface needs establishing between the social media analyzer and a database containing criminal history information. Finally, an interface between the confidence-scoring tool and the criminal history database provides the data needed to compute a confidence score. The product function section explains these interfaces in more depth.

2.2 Product Functions

This software incorporates a number of different applications to perform its function. These applications may be commercially available but may require additional programming to make each work with the confidence scoring application. Each of the following subsections explains the separate software in more detail and describes how each part functions within the final system.

2.2.1 Social Media Analysis Software

Social media analysis software must utilize an API to Twitter and Facebook.

Social media analysis software interfaces with Facebook and Twitter. An API allows the social media analyzer to access content from these sites. The established connections let the analyzer scan Twitter and Facebook posts quickly. Keywords help identify concerning posts. Common keyword searches by investigators include gun, shoot, kill, hit, run over, police, gangs, or other words derived from specific investigations. Past language of posts that ended in violence and recommendations from law enforcement experts' help determine what words accurately identify threats. Lexis Nexis is one company that already uses software to scan Twitter and Facebook for keywords and phrases.[78] Exploring commercially available software like Lexis Nexis can guide developers on this software function or provide a platform of established code.

2.2.2 Language Sentiment

A natural language analyzer must scan each line of text input from the social media analysis software. The natural language analyzer must function within the social media analyzer software.

Natural language analyzer software scans each line of text and provides a negative or positive sentiment.[79] The analyzer does this by identifying proper and common nouns

[78] "Defense Community and Homeland Security," Lexis Nexis Special Services, Inc., July 8, 2017. http://lexisnexisspecialservices.com/who-we-serve/defense-department/.

[79] "Natural Language API Basics," Google Cloud Platform, accessed July 8, 2017, https://cloud.google.com/natural-language/docs/basics.

within text and providing a negative, positive, or neutral emotion toward the noun by the poster.[80] This means the software can read " I hate the Cowboys" and return an opinion that the writer has a negative opinion toward the Cowboys. Taking it even further, language sentiment software can capture the context of messages.[81] A message that reads "My flight was cancelled. Great!" could be read as positive.[82] The software can recognize that "Great!" in this instance is negative.[83] The social media analyzer and the sentiment tool must be designed to work together and compute a confidence value that passes to the confidence-scoring tool.

2.2.3 Public and Criminal Record Search

Criminal record information from public records or a RMS must be accessible through an SQL server or another appropriate interface.

Public record databases house identifying information pulled from motor vehicle records, criminal convictions, or judgements and liens. Running the owner of the social media account through this database returns information on his place of residence, his associates, and his criminal convictions. Lexis Nexis also provides a way to pull local law enforcement RMS and CAD information, allowing a comprehensive score for a poster's criminal background.[84] If a person has a criminal background, this part of the software assigns a score based on the conviction crime type. For example, if the poster is John Smith, the software looks for other identifying information, such as hometown or date of birth, on the social media account. For instance, this information may show that John Smith has a birthdate of January 2, 1973, lives in Experiment, Georgia, and has a criminal conviction for communication of threats and assault. Since this identity is tied with a post that has a negative sentiment, this person's criminal conviction is compared to

[80] "Natural Language API Basics."

[81] Kristian Bannister, "Understanding Sentiment Analysis: What it is & Why it is Used," Brandwatch, January 26, 2015, https://www.brandwatch.com/blog/understanding-sentiment-analysis/.

[82] Bannister.

[83] Bannister.

[84] "Accurint Crime Analysis Workstation," Lexis Nexis Risk Solutions, accessed July 8, 2017, http://www.lexisnexis.com/risk/products/government/accurint-crime-analysis-workstation.aspx.

a predetermined list. Certain crimes score higher than others on this list. Hypothetically, the communication of threats charge may receive a score of three while the assault may score four points. Crimes that involve personal injury score higher than crimes involving property damage. The score passes to the next stage in the software, confidence scoring.

2.2.4 Confidence Scoring

The language sentiment value and the criminal history data value interface with the confidence-scoring algorithm. The confidence score must be calculated for a confidence threshold. When the confidence threshold is met, an alert signals the user.

Confidence scores are numbers assigned to the certainty for findings.[85] As the name implies, the scores can show a level of probability an event might happen.[86] In the case of threat investigations, alerting investigators when a threshold is reached is more important than finding the probability. This threshold is a predetermined score that if reached, requires assignment of the case to an investigator. A critical step in this process is determining the correct threshold to cause that alert. This SRS follows the below steps proposed by Bill Murdock for determining these thresholds:

1. Assign numerical rewards to each possible outcome
2. Run a large number of queries for which you know the outcomes.
3. For each possible threshold, compute the net reward for the system at each possible threshold.
4. Select the threshold that has the greatest reward.[87]

This process involves using threat cases from the past, assigning values for the language sentiment and criminal history components as well as running the queries through the system.

One way to obtain these values and calculate confidence scoring is to calculate trust scores.[88] If an investigator performs confidence scoring manually, he has to get the

[85] Bill Murdock, “How to select a threshold for Acting Using Confidence Scores.”

[86] Murdock.

[87] Murdock.

[88]“How to Calculate a Confidence Score,” CrowdFlower, accessed July 7, 2017, https://success.crowdflower.com/hc/en-us/articles/201855939-How-to-Calculate-a-Confidence-Score.

help of a number of experienced investigators. Each investigator rates each post's language sentiment and criminal history as well as assigns a trust value to each. These data are then sent through the following series of calculations:

1. Sum the trust scores of the contributors for each response.
2. Sum the trust scores of all the contributors
3. Divide each in (1) by (2) to find the confidence score for each response.[89]

The results from calculating trust scores for a series of threatening posts are compared against threat follow-through to violence. A trust score that corresponds with an outcome of violence is picked for an alert to investigators for follow-up. Manually computed threat scores provide the baseline for the software to calculate the thresholds. The performance of these thresholds may change over time as machine-learning identifies changes in the threat scenarios, so threshold calibration is evaluated routinely.

2.3 User Characteristics

Law Enforcement makes up the user of this software. The investigators' function includes threat investigations but also includes handling multiple cases of different crime type. Multiple investigators may work threat cases. They cannot devote all their time monitoring software for an alert. Time at a desk varies from case to case. This group of user usually has a smartphone and is familiar with using applications on it. Important software characteristics for this user are the ability to receive alerts on multiple computers and/or smartphones in one agency. Online access to the software report also benefits the user. They can access the report anywhere and start interventions as needed. Printable reports help preserve evidence and complete case files.

2.4 Constraints

Since this is law enforcement software, certain boundaries must be maintained in accessing citizen's information. Most of these constraints center on maintaining the privacy of people who do not post threatening language. The first constraint is that he law

[89] "How to Calculate a Confidence Score."

enforcement user cannot see the social media language until after an alert. The second is that law enforcement cannot access the algorithms used in the software. The third is that the software cannot store social media posts that do not alert investigators for more than 90 days. Finally, the software must be able to store data related to alerts indefinitely. Cases that result in charges can last for years and evidence must be kept. In the case of homicides, even after trial, the evidence must be kept in case of appeal for many years. Each jurisdiction's rules are different so the ability to retain certain data for different lengths of time is important. Rules about evidence storage apply to this constraint as they apply to each individual case.

3. CASE STORIES

In the previous chapter on manual investigations, four case stories explored the way investigator subjectivity influences cases. In this user case section, the same incidents help illustrate how the proposed software guides investigators in determining which cases to give priority. This section looks at the same cases and applies language sentiment, criminal history, and confidence scoring in order to evaluate threat risk. Normally user cases illustrate how the user interfaces with the software. Since after initial setup this program runs independent of the user, the cases here demonstrate the pathway the software uses in determining whether a threat reaches the confidence threshold. In order to accomplish this task, some assumptions need defined. The language sentiment tool uses a hypothetical value range of 0–10. The criminal history section assigns values within the range of 0–5. The confidence scoring range assigns values of 0–10. The hypothetical threshold value for these studies is 5. The values representing calculations from sentiment analysis and criminal histories were determined by the author's 23 years of experience investigating threats, not by a group of experts as normally is necessary to achieve more accurate ratings.

3.1 Facebook Post against Officer Twiddy

In February 2016, a Raleigh Police officer shot a suspect during a struggle. Shortly after the incident, people from the community started posting on Facebook and

Twitter. A lot of these posts talked negatively about the police. Other posts communicated threats toward law enforcement. This case study is taken from a Facebook post made after the shooting. The person posting was identified and referred to here as B. K. He posts the following comment:

> Im sorry I don't normally think about this but peaceful protesting? For what? For them to continue to take us African Americans as a joke. I may be wrong but the way I'm feeling it's time to shoot some of they kind down too. We always talk about other cities going through what we going through && what we would do. I'm going to support my city 100% RIP Lockman[90]

This post is followed by a comment from another subject wanting peaceful protests in his neighborhood, fearing houses my get shot up where they live. B. K. replies to this individual with

> Ain't nobody thinking about shooting up nobodies house or kids…….I said shoot some of their kind down not go shoot up houses and kill kids.[91]

B. K. then posts on his Facebook page a picture of two t-shirts, one reading "F*** a cop named Twiddy" and the other "100 to 500 RIP Lock AKA."[92]

3.1.1 Language Sentiment

Preprogrammed keywords locate the post's language of "shoot some of they down," and "F*** a cop."[93] Each line of the poster's account is scanned by the language sentiment tool. B. K.'s post is very negative. He questions the usefulness of peaceful protests and suggests the police do not respect African Americans. He also says he feels it is time to shoot a cop. The addition of the word *feeling* is the only thing keeping this post from being a direct threat. Each sentence receives a positive, negative or neutral score, which is then assigned a number that correlates with a degree. Table 1 shows the sentence and the assigned value for negative sentiment toward police.

[90] B.K. Facebook page, https://facebook.com/prfile.php?id=100010733497326&fref=ts.

[91] B.K. Facebook page, https://facebook.com/prfile.php?id=100010733497326&fref=ts.

[92] B.K. Facebook page, https://facebook.com/prfile.php?id=100010733497326&fref=ts.

[93] B.K. Facebook page, https://facebook.com/prfile.php?id=100010733497326&fref=ts.

Table 1. B. K.'s Post Sentiment[94]

Social Media Text	Sentiment Value
Im sorry I don't normally think about this but peaceful protesting	3
For what	0
For them to continue to take us African Americans as a joke.	3
I may be wrong but the way I'm feeling it's time to shoot some of they kind down too.	7
We always talk about other cities going through what we going through && what we would do.	2
I'm going to support my city 100% RIP Lockman	2
Ain't nobody thinking about shooting up nobodies house or kids	0
I said shoot some of their kind down not go shoot up houses and kill kids.	7
F*** a cop named Twiddy	7
100 to 500 RIP Lock AKA	2

The average of these values is 3.3 and passes to the confidence scoring tool.

3.1.2 Criminal Background

Once the language sentiment analysis is complete, the poster's identity moves into a search engine that researches criminal backgrounds through a public records database or tied into an agency database. B. K. has no violent crime, harassment or offenses of communicating threats in his criminal history. A specific criminal history for B. K. is unavailable since the investigation into his posts does not result in criminal charges. For

[94] B.K. Facebook page, https://facebook.com/prfile.php?id=100010733497326&fref=ts.

this study, a value of 0 is used since nothing violent or threat-related has been documented.

3.1.3 Confidence Scoring

The language sentiment score and the criminal history score move to the confidence scoring process. In B. K.'s example, the 3.3 from the sentiment score and the 0 from the criminal history score automatically proceed to the confidence scoring algorithm. For these case studies, the confidence score is the sentiment score added to the criminal history score. The calculated value of 3.3 does not meet the predetermined threshold of 5 so the software does not alert investigators.

3.2 Facebook Post against the Raleigh Police Department

This second case study looks at a post following the same incident as the first case study and explores a series of posts made by T. W. spanning March 3 to March 6, 2016. T. W.'s identity was known. On Facebook, T. W. writes,

> […] if RPD [Raleigh Police Department] lies one more time, I'm setting the whole shit on fire tonight." Several comments to this post tried to dissuade T.W. from using fire. T. W. replied with "… assembling today to show force. Meet at 5pm. […][95]

T. W.'s next post is

> I can't continue to be docile. .. I am ready to ride or die (bomb, gun, bomb [emoji]) I WILL GO ALONE IF NEED BE.[96]

Three people post comments telling him not to resort to violence. T. W. writes, "They're leaving us no choice," which received the comment from another poster "we always have choices."[97] T. W. responds, "You have to be willing to die for what you believe in."[98]

[95] T.W. Facebook page, https://facebook.com/mr.nething.

[96] T.W. Facebook page, https://facebook.com/mr.nething.

[97] T.W. Facebook page, https://facebook.com/mr.nething.

[98] T.W. Facebook page, https://facebook.com/mr.nething.

3.2.1 Social Media Monitoring and Language Sentiment

Programmed keywords locate the RPD reference in this post along with "I'm setting the whole shit on fire tonight."[99] Scanning his posts sentence by sentence produces Table 2 of sentiment values.

Table 2. T.W. Post Sentiment[100]

Social Media Text	Sentiment Value
if RPD [Raleigh Police Department] lies one more time, I'm setting the whole shit on fire tonight	7
assembling today to show force. Meet at 5pm	5
I can't continue to be docile	5
I am ready to ride or die (bomb, gun, bomb [emoji]) I WILL GO ALONE IF NEED BE.	4
They're leaving us no choice	4
You have to be willing to die for what you believe in	4

This post has concerning language in it with the poster saying he is "setting the whole shit on fire tonight."[101] The poster adds a criterion that RPD has to lie before he acts. This additional statement lessens the likelihood of violence. T. W.'s entire text is very negative, but the other sentences do not contain further threats. The sentiment score for T. W. is a 4.8.

[99] T.W. Facebook page, https://facebook.com/mr.nething.

[100] T.W. Facebook page, https://facebook.com/mr.nething.

[101] T.W. Facebook page, https://facebook.com/mr.nething.

3.2.2 Criminal Background

The program then looks at T. W.’s criminal history. Running his name through the public search engine portion of the software reveals no violent criminal history, no harassment, and no communication of threat. This information receives a score of 0, and the score moves into the confidence scoring section.

3.2.3 Confidence Scoring

The language sentiment score and the criminal history score move to the confidence scoring section. In T. W.’s example, the 4.8 from the sentiment score and the 0 from the criminal history score input into the confidence scoring algorithm. The confidence score is the sentiment score added to the criminal history score. In this example, the algorithm returns a confidence score of 4.8. The threshold is predetermined as 5, so although this case is borderline, it does not alert investigators.

3.3 Facebook Post to the Republican Party

James T. Hodgkinson shot members of the Republican Party as they practiced baseball on June 14, 2017. Law enforcement shot and killed Hodgkinson during the attack. This case study examines his Facebook posts. This study starts with a post on March 22, 2017, that reads

> Trump is a Traitor. Trump Has Destroyed Our Democracy. It’s Time to Destroy Trump & Co.[102]

The posts on Facebook surrounding this one center around the poster’s hate for the Republican Party. Other posts include language such as

> I Want to Say Mr. President, for being an a** hole you are Truly the Biggest A** Hole We Have Ever Had in the Oval Office,

posted on June 12, 2017, and

> Republican B**ch [a reference to Republican Karin Handel on her comment that she doesn’t support a living wage] Wants People to Work

[102] “Saved From https://www.facebook.com/jthodgkinson.”

> for Slave Wages, when a Livable Wage is the Only Way to Go! Vote Blue, It’s Right for You!

posted on June 8, 2017.[103]

3.3.1 Social Media Monitoring and Language Sentiment

This case does not contain clear-cut threatening language. The language in the March 22, 2017, post says that “It’s Time to Destroy Trump & Co.,” but *destroyed* is not a common keyword for violent acts.[104] It is possible that the Secret Service, in its duty to protect the president, would monitor for this broader term. Running the program for negative sentiment against Trump should flag these posts. The sentiment of the language is negative but does not indicate direct violence.

The first clearly negative post toward Trump came in March 2017. Around 26 other accessible posts with multiple sentences are available for the time span from March 2017 through the shooting on June 14, 2017. Table 3 shows the assessable comments and the hypothetical scoring.

[103] “Saved From https://www.facebook.com/james.hodgkinson.568.”

[104] “Saved From https://www.facebook.com/jthodgkinson.”

Table 3. Hodgkinson Post Sentiment[1]

Social Media Text	Date	Sentiment
Trump is a Traitor. Trump Has Destroyed Our Democracy. It’s Time to Destroy Trump & Co.	3/22/2017	5
Sign the petition for independent investigation	5/21/2017	0
Sign the petition to stop the NEXUS pipeline	5/24/2017	0
Emperor Maximus Imbecilus repost picture of trump	6/3/2017	3
Gov. Jerry Brown Tells Trump He’s Wrong .... What Else Is New?	6/3/2017	3
Willie is a Good Dude. (willie picture saying we need to treat others fairly)	6/4/2017	0
Share if you agree on limiting congress to 2 terms like the president	6/4/2017	0
A lot of guessing (picture of science students)	6/5/2017	0
Everyone Must Register to Vote & When the Day Comes You Must Vote	6/5/2017	0
Lame! (how well do you know your hippie slang quiz)	6/5/2017	0
Trump flunks most tests (reference headline of Trump flunking leadership test after London attack)	6/5/2017	3
Coincidence? I Think Not! (reference to Trump speech during Comey hearing)	6/8/2017	4
Founding Fathers Would Hate What Our Democracy Has Morphed Into. We Now Have an Aristocracy, a Corporatocracy, & an Oligarchy, & a Plutocracy. This Turns Our Country Into a Fascist State! Vote Blue, It’s Right for You!	6/8/2017	5
Republican Bitch Wants People to Work for Slave Wages, when a Livable Wage is the Only Way to Go! Vote Blue, It’s Right for You! (about a headline of Georgia republican not supporting livable wage.	6/8/2017	6

[1] “Saved From https://www.facebook.com/jthodgkinson,”; and “Saved From https://www.facebook.com/james.hodgkinson.568.”

Social Media Text	Date	Sentiment
That should do it (reference picture with caption How to solve global warming. Convince republicans that rising temperatures are turning people gay.)	6/8/2017	6
Shared memory of Sanders campaign	6/10/2017	2
Please. (Image with text: stop fighting over who created the world and fight against the people who are destroying it.)	6/11/2017	1
Bernie Sanders Day post	6/11/2017	1
Trump is a Mean, Disgusting Person.	6/12/2017	6
Repeat of Bernie Sanders day	6/12/2017	1
Trump is Guilty & Should Go to Prison for Treason.	6/12/2017	7
Are You One of the Twenty-Three Million? That’s a Lot of People	6/12/2017	0
Make America Great Again, Resign! (pic: Trump with all n all just another prick in the wall)	6/12/2017	6
I Want to Say Mr. President, for being an ass hole you are Truly the Biggest Ass Hole We Have Ever Had in the Oval Office.	6/12/2017	7
Closed Primaries are What Third World Countries Have. Open them for all Candidates.	6/12/2017	4
That’s Exactly How It Works.....(pic with How does a bill work? Corporation writes bill and then bribe congress until it becomes law)	6/13/2017	6

Assigning each piece of text a sentiment value and calculating an overall score produces a score of 2.9. This may seem like a low score for the amount of anti-Trump, anti-Republican, and anti-government language. It is not uncommon for people to voice their negative opinions on topics they feel strongly about, so these post, while negative, are not threatening.

3.3.2 Criminal Background

The account owner is clearly identified as James T. Hodgkinson. The software searches criminal history information on his identity. Substituting for the software in this instance, news reports provide information based on public record searches: James Hodgkinson has prior arrests for two assaults and an aggravated discharge of a firearm.[1] The charges did not end in convictions. The assaults and firearms charges are violent crimes and produce a score that averages all three charges to a 4.

3.3.3 Confidence Scoring

The language sentiment score and the criminal history score move to the confidence scoring section. In Hodgkinson's example, the 2.9 from the sentiment score and the 4 from the criminal history score input into the confidence scoring algorithm. The addition of the two numbers returns a confidence value of 6.9. The threshold is predetermined as 5, so this case causes an alert to investigators that they need to investigate Hodgkinson further.

[1] Jose Pagliery, "Suspect in Congressional Shooting was a Bernie Sanders Supporter, Strongly Anti-Trump," CNN, June 15, 2017, http://www.cnn.com/2017/06/14/homepage2/james-hodgkinson-profile/index.html.

3.4. Facebook Post to New York Police Officers

On July 5, 2017, John Bonds shot and killed a New York city police officer as she sat in a command vehicle. Bonds posted a Facebook Live video that contains language against the police. As with the other three case studies, this study looks at the posts through its language, the poster's criminal history, and confidence scoring.

The other studies contained language in written posts. This case study involves language posted to Facebook through a video that was transmitted live and then saved as a post on the site. The *Wall Street Journal* reports that Bonds made the video around September 2016 and it contained the following language:

> I'm not playing, Mister Officer. I don't care about a hundred police watching this s—shit. You see this face. You see this face or anything, leave it alone. Trust and believe, [and]
>
> I'm not hesitating. It ain't happening, Bond added. I wasn't a b—- bitch in jail, and I'm not going to be a b—- bitch in the streets.[2]

Other comments from this video are documented by the *Daily News* as

> Y'all n-----s so reluctant to want to say something to the police, man, Man, police is f----ts, and this ain't no gimmick. F----ts. [and] N-----s ain't taking it no more, Mr. Officer. I'm here to tell you, man. ... just keep your a-- away from mine.3

3.4.1 Social Media Monitoring and Language Sentiment

While monitoring for keywords, the software, which would need to have the ability to translate speech, picks up the word *police* in Bond's video. The whole video is very negative toward the police but does not contain any direct threats. It only commands police to leave Bonds alone. The software evaluates the text and produces a score. Table 4 shows the sentence-by-sentence scoring, which is used to determine the overall score for the category.

[2] Phillips, Berman and Lowery, "I'm Not Playing Mr. Officer."

[3] Adam Shrier, Graham Rayman, and Larry Mcshane, "NYPD Cop Killer Alexander Bonds Posted Anti-Police Facebook Rant," *Daily News*, July 5, 2017, http://www.nydailynews.com/new-york/nyc-crime/suspected-nypd-shooter-assaulted-officer-brass-knuckles-article-1.3302356.

Table 4. Bond Post Sentiment[4]

Social Media Text	Sentiment Value
I’m not playing, Mister Officer	2
I don’t care about a hundred police watching this s—shit.	2
You see this face	1
You see this face or anything, leave it alone	3
Trust and believe	0
I’m not hesitating	1
It ain’t happening	1
I wasn’t a b—- bitch in jail, and I’m not going to be a b—- bitch in the streets	3
Y’all n-----s so reluctant to want to say something to the police, man, Man, police is f----ts, and this ain’t no gimmick	3
F----ts	3
N-----s ain’t taking it no more, Mr. Officer. I’m here to tell you, man	3
just keep your a-- away from mine	4

Since the post is located after an act of violence 10 months later, this study assumes the video is located shortly after its posting. Under these assumptions, a computed sentiment value returns a 2.2. This score passes to the confidence scoring software.

[4] Shrier, Rayman, and Mcshane, “NYPD Cop Killer Alexander Bonds Posted Anti-Police Facebook Rant,”; and Phillips, Berman, and Lowery, “I’m Not Playing, Mr. Officer.”

The difficulty of this study is that the violence took place 10 months after the video post. According to NBC New York, this video is followed by inspirational posts.[5] The length of time with no other concerning posts may lower the negative sentiment score further as more and more inspirational posts come online. If the post is located soon after being put online, intervention may hopefully prevent this scenario.

3.4.2 Criminal Background

Bonds is easily recognized as the Facebook page owner. When his identity moves through databases for criminal records, his convictions show a robbery with a firearm, a drug violation, and an assault on an officer.[6] Since it is unknown whether local law enforcement had Bond's mental illness in its RMS, the system may or may not locate information on involuntary commitments (IVC). This case assumes that the software does not locate his mental diagnosis. Based on criminal history alone, a value of 6 is returned for Bonds.

3.4.3 Confidence Scoring

The language sentiment score and the criminal history score move to the confidence scoring section. In Bond's example, the 2.2 from the sentiment score and the 6 from the criminal history score are input into the confidence scoring algorithm, returning a score of 8.2 The threshold is predetermined a 5, so this case alerts investigators that they need to investigate further.

[5] Marc Santia, "Who is Alexander Bonds? Gunman Who Killed NYPD Cop Once Ranted Online About Treatment in Prison," NBC New York, June 5, 2017, http://www.nbcnewyork.com/news/local/Alexander-Bonds-Gunman-NYPD-Officer-Shooting-Ambush-Bronx-432666103.html.

[6] Santia.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. IMPLEMENTATION ANALYSIS

Law enforcement officers love new toys, which may take the form of new cars with cutting-edge LED lights, the latest RADAR unit, the fastest computer, or the newest technology software that helps officers solve cases. It may seem unusual for officers to be excited about software, but today's law enforcement, whether assigned to patrol or detective work, uses technology constantly. If the officers perceive that software makes them better at solving their cases, they want to use it. Unfortunately, putting new software to use for law enforcement involves more than officers wanting it. Questions such as who uses the software, how it is used, how it improves job performance, and how it impacts citizens need answering. To show the vetting process prior to implementing a new piece of software, this chapter takes a look at the issue of case decisions made by software, the effects on citizen privacy, and the design obstacles for the threat-analysis software described in Chapter IV.

A. CAN SOFTWARE HELP MAKE CASE DECISIONS?

Software that makes decisions about what cases are investigated seems a bit unrealistic. The idea is not very far from current practice since decisions happen everyday on what gets investigated and what does not. Each jurisdiction is allocated a certain number of officers to investigate crimes. Each officer continues to receive cases most days of the week. Using Raleigh, North Carolina, as an example, a city with a population of around 450,000, close to 35,000 cases come in for investigation annually.[7] There are approximately 100 investigators assigned to the detective division. If this caseload is divided equally among the 100 detectives, each detective receives 350 cases to investigate per year, each case requiring different amounts of time and resources. With this high case load, detectives must decide to which cases they dedicate the most effort Currently, investigators employ solvability factors to determine those cases. While these

[7] "Quickfacts," United States Census Bureau, accessed July 8, 2017, https://www.census.gov/quickfacts/table/PST045216/3755000; and Raleigh Police Intelligence Center, unpublished data, April 20, 2017.

solvability factors currently influence case investigations, the advent of big-data opens the door to software making investigative decisions. To explore the question of whether this concept is feasible, this section discusses the current use of solvability factors and then data-driven decisions.

Solvability factors are information and evidence that is available to solve a case.[8] Urlacher and Duffy provide the following 12 factors, one of which must be present to solve a crime:

- witnesses to the crime,
- knowledge of the suspect's name,
- knowledge of where the suspect can be located,
- description of suspect, identification of suspect,
- property with traceable, identifiable characteristics, marks or numbers,
- existence of a significant method of operation,
- presence of significant physical evidence,
- description of the suspect's vehicle,
- positive results from a crime scene evidence search,
- belief that crime may be solved with publicity
- reasonable additional investigative effort,
- possibility and/or opportunity for anyone, other than the suspect, to have committed the crime.[9]

The more factors investigators have, the higher the possibility of solving the case. Cases with a high probability of being solved are assigned for investigation. Cases that have no solvability factors are not investigated. While not investigating a case seems callus, an investigator's time is better spent working on cases that have a greater likelihood of ending in an arrest. All types of crime are reviewed for their solvability. Even a homicide investigation ends once the solvability factors end. Presently, each officer decides whether a crime happened and whether it is solvable.

Threat software employs this same decision making as it evaluates threats over social media. Using language sentiment and confidence scoring, the software shows which threats are likely to cause violence. This equates to showing which language on

[8] Peter B. Bloch and James Bell, *Managing Investigations: The Rochester System* (The Police Foundation, 1976), 45.

[9] Bloch and Bell., 45.

social media rises to the statutory level of communicating a threat, satisfying the first step in crime investigation—ensuring a crime actually happened. The ability of the software to learn over time increases its accuracy, thus leading to officers bringing charges for cases that have a high probability of ending in violence. This accuracy also helps officers determine their solvability factors.

One of those factors is the "reasonable additional investigatory effort."[10] For example, in a homicide case, a department may find it reasonable to spend several thousand dollars for DNA analysis, while the same department may decide that spending the same money for DNA analysis on a property crime investigation is unreasonable. Communications of threat crimes are misdemeanors in most jurisdictions and receive lower investigative priority. Investigating threats over social media requires additional techniques such as obtaining court orders and interpreting IP information, conflicting with the reasonable effort for the crime solvability factor. Software identifying the likely threats based on data of past violence increases the seriousness of the threat and justifies putting in the investigative effort. The Hodgkinson and Bond case studies in the previous chapters show how manual, subjective application of solvability factors may miss posts that end in violence. When the software assists in locating the threats, its alerts may allow early intervention.

The process of the software may be compared to the current practice of allowing computers to make decisions for patrol allocation. Predictive policing software, such as Predpol, is an example of agencies accepting the direction of resources by computer software.[11] Software, such as Predpol, analyzes data contained in department RMS and CAD systems to determine, through proprietary algorithms, areas where officers need to concentrate policing efforts to prevent crime.[12] These programs run automatically after software administrators determine the period and certain crimes they want the formulas to consider. Results show on a map as a square around a geographic area. Police supervisors direct resources to these areas, possibly leaving other areas untouched or with less police

[10] Bloch and Bell, 45.

[11] "Predpole is Predictive Policing," Predpole, accessed July 8, 2017, http://www.predpol.com/about/.

[12] "Predpole is Predictive Policing."

presence. The computer directing officers to patrol geographic regions equates to a computer telling a detective which threats to investigate. Machine learning lets the software refine its algorithms with better language and criminal history combinations that result in violence. An accurate threshold for the alert lets the software allocate only those threats with the highest potential for investigation.

Allocating only the highest risk threats for investigation is important to agencies with limited resources, but not investigating threats is problematic as well. Threats usually communicate a violent act toward an individual. Deciding not to investigate one of these threats leaves the possibility of having to answer for why it was not investigated if violence does happen. Saying “the machine did not alert to that one” will not go over well with a victim’s family, leading to negative publicity and possible job repercussions. This is not a new problem as citizens already complain about how their individual cases are handled. Understanding how the software makes its recommendations and showing that its reliability is equal to or greater than a detective’s manual decisions are important for this reason.

As illustrated in the preceding paragraphs, threat software is in line with current investigative practices and current uses of technology. Computers already help make deployment and investigative decisions for departments. This new software is consistent with current practices in that it searches for connections between people in hopes of identifying suspects, their associates, and related criminal histories to increase the accuracy of the confidence in risk of occurrence. These steps are now performed manually with the aid of online database search engines and with the confidence scoring done subjectively by investigators. The software reduces the subjectivity by using historical data to predict outcomes. As more data are input into the software, the accuracy and reliability should increase. Providing a consistent, reliable method for these investigations can ensure each is treated equally, reducing the claim that a certain threat was not considered. The claim is easily answered since the software considers all language but moves on to new posts once the last one is ruled out for a risk of violence.

Now that the discussions show the software is in line with current practice, is the use of big data through data-driven decisions beneficial? Data-driven decisions are decisions based on information instead of using just your experience.[13] Software lets these decisions happen by an automatic process, removing human subjectivity. Data collected over time helps predict future events. An example of this, as highlighted by Provost and Fawcett, is Walmart's use of data from a previous hurricane to predict what items to sell before an upcoming storm makes landfall.[14] In this case, the data predict human behavior. Companies that are leaders in their industries make 6 percent higher profits when they use big data to predict human behavior.[15] McAfee and Bryonjolfsson make a correlation between these high data-driven decision companies to the fact each places the decision makers where they have access to the data.[16] Placing new software in case management decisions helps enhance decision making on threat case. The use of big data takes the intuition out of the investigator's decision process. Instead, it uses the history of human behavior to guide investigative decisions.

B. DOES THE SOFTWARE CHANGE CITIZEN INTERACTIONS?

Social media and law enforcement conjure up an image of the National Security Agency collecting data from everyone's Facebook pages and Twitter accounts. The perception of government collecting citizen data is unpopular. Companies that have provided services to the government to scan social media, such as Geofeedia, have lost their contracts with Facebook and Twitter when citizens learned how their information was being used.[17] Losing contracts does not mean law enforcement did anything illegal with the information. Facebook's and Twitter's decisions seem to be based on the fear of public outcry rather than law enforcement misusing the data. Since the topic of social media use by law enforcement is sensitive, any new software has to consider citizen

[13] Foster Provost and Tom Fawcett, "Data Science and its Relationship to Big Data and Data-Driven Decision Making," *Big Data* 1, no.1 (March 2013): 53, doi: 10.1089/big.2013.1508.

[14] Provost and Fawcett, 52.

[15] Andrew McAfee and Erik Brynjolfsson, "Big Data: The Management Revolution," *Harvard Business Review*, (October 2012): 6, http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf.

[16] McAfee and Brynjolfsson, 8.

[17] Lora Kolodny, "Facebook, Twitter Cut Off Data Access for Geofeedia, a Social Media Surveillance Startup," Crunch Network, October 11, 2016, https://techcrunch.com/2016/10/11/facebook-twitter-cut-off-data-access-for-geofeedia-a-social-media-surveillance-startup/.

privacy. Even though the use of social media information is legal, a positive perception surrounding the information use is essential. This section discusses how the threat software follows current legal use of both social media and personal information and helps safeguard how the data are handled.

The threat software incorporates tools already legally used by law enforcement. Law enforcement uses social analytic tools on a routine basis to analyze posts placed for public viewing. The analytic tools only speed up what an individual officer is able to do through manually viewing individual posts. The software allows more posts to be analyzed, but its use cuts down on the number of sites actually viewed by law enforcement officers.

The use of keywords and phrases allows the software to analyze social media language without law enforcement viewing each post. The software only returns those posts that contain the requested information. Looking at the case studies as examples, officers keyed in words such as “kill,” “gun,” and “police.” These words would have returned B.K.’s and T.W.’s post. The use of these terms in a threat provides law enforcement reasonable suspicion to investigate the post for a crime. Identifying only posts with threatening language leaves the citizen’s expectation of privacy intact.

Identifying posts containing threatening language quicker also gives law enforcement intervention options other than arrest. Officers often talk to individuals dealing with life crisis. Often this is enough to prevent escalation. The other benefit is the officer’s interactions may uncover other influences on the person. One of these influences is mental crisis. Officers can work with the poster, the poster’s family and the court system to get the help needed to extract the person from the crisis.

Another area of concern with social media analytic tools is the collecting and retaining of data files on innocent citizens. This is the complaint that citizens voiced after Snowden revealed how the NSA was using cell phone data.[18] The threat software concept includes coding that scans Facebook and Twitter posts but does not provide those posts to

[18] Ewen Macaskill and Gabriel Dance, “NSA Files: Decoded,” *The Gardian* (November 1, 2013) https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded#section/1.

law enforcement until the confidence threshold is triggered. At that point, the completed "package" of information that the software uses in its decision is made available to the investigator. The method would mean that only individuals whose posts rise to the level of reasonable suspicion are investigated. The program is coded to purge data on a routine basis to avoid collecting and retaining information on citizens not suspected of a criminal offense. This purge should not happen as soon as a post is viewed. This would mean missing posts such as Hodgkinson's social media thread wherein a single post does not raise concern, but the continuation and escalation of language does. Purges based on a specified period help prevent law enforcement from missing concerning threads while reassuring citizens that a data file is not kept. Explaining the purging process and showing when it is scheduled may further ease concerns as citizens understand the benefit. This type of policy aligns with the way some law enforcement agencies are regulating their surveillance camera data. Raleigh, North Carolina, has a policy requiring that videos be purged every 90 days.[19]

C. CAN THIS SOFTWARE REALLY BE DESIGNED AND IMPLEMENTED?

Threat software identifies social media posts that pose the highest risk of violent results would benefit the law enforcement community. The goal of law enforcement is to protect life and property, so the hope is that locating more threats means there are fewer victims. The second benefit is higher efficiency for investigators while the third is maintaining citizen privacy. Implementing software to meet these three goals should save lives and increase law enforcement efficiency, but design and implementation obstacles do exist. Therefore, this section discusses how software calibration issues and information sharing need overcoming before a reliable product is released.

Threat software needs historical and real-time data to accurately determine the combination of language sentiment and criminal history likely to end in violence. This data needs to include both threats that end in violence and those that never do. Fortunately for threat victims but unfortunately for the needs of the software, most threats

[19] Raleigh Police Department, *Department Directives: Automated License Plate Recognition and Internet Protocol Camera Systems, 1110–07* (Raleigh, NC: Raleigh Police Department, 2016).

do not end in violence. As seen in the Raleigh incident in 2016, which contained case stories one and two, out of the multiple threats, none ended in violence. If the hypothetical software were to analyze data only from Raleigh, it could not adequately determine the risk for violence. Since most threats over social media in the Raleigh area do not end in violence, it would likely take years to obtain enough data to achieve reliability. This means the software needs access to data from multiple jurisdictions. The ability to obtain enough data is a problem of information sharing.

Information sharing in law enforcement has gotten better since 9/11, but there is no consistent mechanism to share RMS and other case data. A lot of proximate agencies meet regularly and share crime information face to face. Common suspect information passes from one jurisdiction to another during these meetings. Copies of reports get handed over as paper copies. One reason for these meetings is each agency uses a different RMS, each with its own unique fields and none designed to interface with another RMS. As an example, the Raleigh Police Department uses Keystone while the Wake County Sheriff's office uses Sungard OSSI. There is no way to share data between the agencies except by hand-delivery or email. This just shows the difficulty in proximate agencies sharing data. If a nationwide sharing system is needed to obtain enough data to calibrate the software, the problem is obviously much larger. This problem is not too big to overcome. Although individual agencies are not developing ways to combine their data, private companies are. Lexis Nexis recently launched a service that allows agencies a secure way to share their data and run analytics across different agencies. Each organization has to choose whether it wants to pay to enter the service, so it is not a total solution but does show the ability to incorporate individual RMS into one database. Further research can show just how extensive information sharing needs to be to calibrate the software properly.

After establishing a large enough data stream, calibration becomes the next obstacle. As seen in the case studies, the language with the most violent references but from individuals with no criminal histories did not end in violence while the language with the least violent references but offenders with a criminal history did end in violence. Because there is the potential for many different combinations of language and criminal history, establishing the initial calibration is difficult. Machine learning can adjust to "training data," so if enough data is obtained, this obstacle may be insignificant.[20] This data should contain enough historical information to capture as many different combinations of language, criminal histories, and outcomes as possible.

D. CONCLUSION

> *June 12, 2017, started out normal enough for the Secret Service. Individual investigators went about working their cases. Some sat at their desks researching on computers, some went out to interview people involved in their cases, and others took calls from people concerned about cases or reporting suspicious people. Out of sight of the hustle and bustle of the office, a computer churned away, scanning social media feeds for threats against the president. Suddenly, the screen of the lonely computer came to life with a red, flashing banner. Investigators' phones buzzed and beeped simultaneously. Everyone stopped what they were doing to see what important alert had come through. The social media threat assessment tool alerted to a potential threat against Republican Party members.*
>
> *A report on the threat provides Special Agent Smith the details on the case. Facebook posts starting back in March 2017 consistently escalated negative language about President Trump and the Republican Party. The person posting had previously been charged with a violent crime in another state. While Hodgkinson has not directly threatened anyone, the software's calculations setting off the alerts provide evidence that violence happens after people start posting similarly. SA Smith knows he can intervene by taking this case seriously.*
>
> *The investigative work of the case leads SA Smith to find that Hodgkinson is in the Washington, D.C., area. After partnering with local law enforcement, Hodgkinson is located at a local YMCA. SA Smith decides the best course of action is to interview Hodgkinson about his Facebook posts. During this interview, Hodgskinson presents signs of mental deterioration. SA Smith recognizes these signs and with the help of the rapport he establishes with Hodgkinson, gets resources lined up to help. Following up on the case over the next week finds Hodgkinson responding well, and he even returns home to Illinois. The violence that could have played out on June 14, 2017, never happens, and Hodgkinson continues to be an important part of his family.*

[20] Bill Murdock, "How to Select a Threshold for Acting Using Confidence Scores."

As seen earlier in the paper, this event ended with congressmen wounded and Hodgkinson dead. There is no guarantee that a machine or a human could have prevented the shooting on June 14, 2017. However, the proposed social media threat assessment tool has the potential to help law enforcement intervene in a case much earlier, protecting both the victim and the suspect from tragic outcomes.

LIST OF REFERENCES

Alpar, Paul, and Daniel Ohliger. "Creation of Risk Profiles of Business Customers from Social Media." *Banking and Information Technology* 16, no. 1 (March 2015): 26–36.

Association of Threat Assessment Professionals. *Risk Assessment Guideline Elements for Violence: Considerations for Assessing the Risk of Future Violent Behavior*. Sacramento, CA: Author, (2006). https://c.ymcdn.com/sites/ www.atapworldwide.org/resource/resmgr/imported/documents/RAGE-V.pdf.

Bannister, Kristian. "Understanding Sentiment Analysis: What It Is and Why It Is Used." Brandwatch. January 26, 2015. https://www.brandwatch.com/blog/understanding-sentiment-analysis/.

Bloch, Peter B., and James Bell. *Managing Investigations: The Rochester System.* Washington: The Police Foundation, 1976, 45.

Borum, Randy, Robert Fein, Bryan Vossekuil, and John Berglund. "Threat Assessment: Defining an Approach to Assessing Risk for Targeted Violence." *Behavioral Sciences & the Law* 17, no. 3 (September 1999): 331–335. doi: 10.1002/(SICI)1099-0798(199907/09)17:33.0.CO;2-G.

Braun, Richard, and Werner Esswein. "Towards a Conceptualization of Corporate Risks in Online Social Networks: a Literature Based Overview of Risks." In *Enterprise Distributed Object Computing Conference (EDOC), 2013 17th IEEE International*, 267–274. IEEE, 2013. doi: 10.1109/EDOC.2013.37.

Brinkley, Mark. "Social Media Risk." *Internal Auditor* 71, no. 2 (April 2014): 68–69.

Calhoun, Frederick S., and Stephen W. Weston. *Concepts and Case Studies in Threat Management* (Boca Raton: CRC Press, 2013), 1. http://www.crcnetbase.com.libproxy.nps.edu/isbn/9781439892183.

———. "Protecting Judicial Officials: Implementing an Effective Threat Management Process." Washington, DC: Bureau of Justice Bulletin, 2006. http://www.ojp.usdoj.gov/bja.

Carpenter, Serena, and Alisa P. Lertpratchya. "Social Media Communicator Roles: A Scale." *Social Media + Society* 2, no. 1 (January–March 2016): doi:10.1177/2056305116632778.

Crowdflower. "How to Calculate a Confidence Score." Accessed July 7, 2017.https://success.crowdflower.com/hc/en-us/articles/201855939-How-to-Calculate-a-Confidence-Score.

Jerry Cao. “A Practical Approach to Functional Specifications Documents.” Studio. accessed September 15, 2017. https://www.uxpin.com/studio/blog/practical-approach-functional-specifications-documents/.

Dijck, José van, and Thomas Poell. “Social Media and the Transformation of Public Space.” *Social Media + Society* 1, no. 2 (2015): 1. doi: 10.1177/2056305115622482.

Fein, Robert A., and Bryan Vossekuil. *Protective Intelligence and Threat Assessment Investigations: A Guide for State and Local Law Enforcement Officials.* Washington, DC: U.S. Department of Justice, 1998. 1–59.

Fein, Robert A., Gwen A. Holden, and Bryan Vossekuil. *Threat Assessment: An Approach to Prevent Targeted Violence.* Washington, DC: U.S. Department of Justice, 1995. https://www.hitacllc.com/HITAC_Resources/ThreatAssessmentApproachtoTargetedViolence.pdf.

Google Cloud Platform. “Natural Language API Basics.” Accessed July 8, 2017. https://cloud.google.com/natural-language/docs/basics.

Handleman, Lori D., and David Lester. “The Content of Suicide Notes from Attempters and Completers.” *Crisis* 28, no. 2 (2007): 102–104. doi:10.1027/0227-5910.28.2.102.

IEEE. *IEEE Std 830–1998 IEEE Recommended Practice for Software Requirements Specifications.* IEEE Computer Society, 1998.

Jung, Yumi, and Emilee Rader. “The Imagined Audience and Privacy Concern on Facebook: Differences between Producers and Consumers.” *Social Media + Society* 2, no. 2 (2016): 1–15. doi:10.1177/2056305116644615.

Kolodny, Lora. “Facebook, Twitter Cut Off Data Access for Geofeedia, a Social Media Surveillance Startup.” Crunch Network. October 11, 2016, https://techcrunch.com/2016/10/11/facebook-twitter-cut-off-data-access-for-geofeedia-a-social-media-surveillance-startup/.

Learning Theories. “Online Disinhibition Effect (Suler).” December 15, 2015. https://www.learning-theories.com/online-disinhibition-effect-suler.html.

Lexis Nexis Special Services. “Defense Community and Homeland Security.” July 8, 2017. http://lexisnexisspecialservices.com/who-we-serve/defense-department/.

Lexis Nexis Risk Solutions. “Accurint Crime Analysis Workstation.” accessed July 8, 2017. http://www.lexisnexis.com/risk/products/government/accurint-crime-analysis-workstation.aspx.

Long,Colleen, and Jennifer Peltz. "Officer's Killer had Ranted About Police Killing and Abusing." *ABC News,* July 5, 2017. http://abcnews.go.com/US/wireStory/female-police-officer-critical-shooting-bronx-48444782.

Masur, Philipp K., and Michael Scharkow. "Disclosure Management on Social Network Sites: Individual Privacy Perceptions and User-Directed Privacy Strategies." *Social Media+ Society* 2, no. 1 (2016): 1–13. doi: 10.1177/2056305116634368.

Macaskill, Ewen, and Gabriel Dance. "NSA Files: Decoded." *The Guardian*, November 1, 2013. https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded#section/1.

McAfee, Andrew, and Erik Brynjolfsson. "Big Data: The Management Revolution." *Harvard Business Review*, (October 2012): 6, http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf.

Murdock, Bill. "How to select a threshold for Acting Using Confidence Scores." IBM Watson. June 23, 2016, https://developer.ibm.com/watson/blog/2016/06/23/how-to-select-a-threshold-for-acting-using-confidence-scores/.

Neben, Rachel V. "Effectiveness of Threat Assessment Models for Lone Terrorists." *Small Wars Journal* 13, no. 7 (August 2015): http://smallwarsjournal.com/jrnl/art/effectiveness-of-threat-assessment-models-for-lone-terror.

Learning Theories. "Online Disinhibition Effect (Suler)." December 15, 2015. https://www.learning-theories.com/online-disinhibition-effect-suler.html.

Pagliery, Jose. "Suspect in Congressional Shooting was a Bernie Sanders Supporter, Strongly Anti- Trump." CNN, June 15, 2017. http://www.cnn.com/2017/06/14/homepage2/james-hodgkinson-profile/index.html.

Patton, Desmond U., Jeffrey Lane, Patrick Leonard, Jamie Macbeth, and Jocelyn R. Smith. "Gang Violence on the Digital Street: Case Study of a South Side Chicago Gang Member's Twitter Communication." *New Media & Society* (January 2016): 1–19. doi: 1461444815625949.

Phillips, Kristen, Mark Berman, and Wesley Lowery. "I'm Not Playing, Mr. Officer': Gunman Appears to Complain About Police Mistreatment in Video Months Before Shooting NYPD Officer." *Washington Post,* July 5, 2017. https://www.washingtonpost.com/news/post-nation/wp/2017/07/05/assassinated-nypd-officer-shot-and-killed-while-sitting-in-a-police-vehicle-officials-say/?utm_term=.a1962cdcfc1.

Provost, Foster, and Tom Fawcett. “Data Science and its Relationship to Big Data and Data-Driven Decision Making,” *Big Data* 1, no.1 (March 2013): 52, doi: 10.1089/big.2013.1508.

Raleigh Police Department, *Department Directives: Automated License Plate Recognition and Internet Protocol Camera Systems, 1110–07*. Raleigh, NC: Raleigh Police Department, 2016.

Rayman, Graham, and Larry Mcshane. “NYPD Cop Killer Alexander Bonds Posted Anti-Police Facebook Rant.” *Daily News*, July 5, 2017. http://www.nydailynews.com/new-york/nyc-crime/suspected-nypd-shooter-assaulted-officer-brass-knuckles-article-1.3302356.

Santia, Marc. “Who is Alexander Bonds? Gunman Who Killed NYPD Cop Once Ranted Online About Treatment in Prison.” NBC New York, June 5, 2017. http://www.nbcnewyork.com/news/local/Alexander-Bonds-Gunman-NYPD-Officer-Shooting-Ambush-Bronx-432666103.html.

Shepherd, Tamara, Alison Harvery, Tim Jordan, Sam Srauy, and Kate Miltner. “Histories of Hating.” *Social Media + Society* 1, no. 2 (July–December 2015): 1–10. doi:10.1177/2056305115603997.

Shrier, Adam, Graham Rayman, and Larry Mcshane. “NYPD Cop Killer Alexander Bonds Posted Anti-Police Facebook Rant.” *Daily News*. July 5, 2017. http://www.nydailynews.com/new-york/nyc-crime/suspected-nypd-shooter-assaulted-officer-brass-knuckles-article-1.3302356.

Sipior, Janice C., Burke T. Ward, and Linda Volonino. “Benefits and Risks of Social Business: Are Companies Considering E-Discovery?.” *Information Systems Management* 31, no. 4 (Fall 2014): 328–339. doi:10.1080/10580530.2014.958031.

Smith, Allison. “From Words to Action: Exploring the Relationship between a Group’s Value References and Its Likelihood of Engaging in Terrorism,” *Studies in Conflict and Terrorism* 27, no. 5 (2004): 409–437. doi:10.1080/10576100490483679.

Stepashkin, M. V., and F. F. Khusnolarov, “Risk Analysis for Reputation Based on Assessments and Ranking of Information Events and Specific Data from Open Sources of Information.” *Problems of Economic Transition* 57, no. 12, (2015): 8–16. doi:10.1080/10611991.2015.1161443.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California