Human Trafficking Network Investigations: The Role of Open Source Intelligence and Large-Scale Data Analytics in Investigating Organized Crime

Leah F. Meyer and Louise I. Shelley[1]

Leah F. Meyer founded the Human Trafficking Intelligence Project to research human trafficking and equip investigators with tactically relevant tools to uncover trafficking operations. She previously worked on DARPA's Memex project as Project Manager and Investigative Analyst, providing investigative support to law enforcement and building software to investigate human trafficking. She worked with the international NGO, Polaris, building the BeFree SMS hotline and serving as Regional Specialist for the National Human Trafficking Hotline, developing case response and building capacity for law enforcement, service providers, and government agencies to serve human trafficking victims. Email: Leah.Meyer@htip.org

Louise Shelley is the Omer L. and Nancy Hirst Endowed Chair and a University Professor at Schar School of Policy and Government, George Mason University. She founded and directs the Terrorism, Transnational Crime and Corruption Center (TraCCC). Her most recent books include *Dark Commerce: How a New Illicit Economy Threatens our Future* (PrincetonUniversity Press, 2018), written while an inaugural Andrew Carnegie Fellow 2015-2017, *Dirty Entanglements: Corruption, Crime and Terrorism* (Cambridge University Press, 2014), and *Human Trafficking: A Global Perspective* (Cambridge 2010). She is the recipient of the Guggenheim, NEH, IREX, Kennan Institute, Fulbright, and Rockefeller Fellowships. Email: Lshelley@gmu.edu

Abstract

This paper examines the migration of trafficking for sexual exploitation to the web and explores open source research techniques, analytical tools, and datasets used to uncover a Chinese organized crime network engaged in human trafficking. Memex, a US government research program that produced a large dataset and software application tools, provided surface and deep web intelligence through escort advertisements and sex buyer review forum posts to law enforcement investigators. The tools provided visualizations

1 The research done for this paper by Louise Shelley was done under the NSF grant: EAGER:ISN: A New Multi-Layered Network Approach for Improving the Detection of Human Trafficking, Award Number: 1837881.

to explore the relationships among seemingly disparate online advertisements using attributes, such as telephone numbers, email addresses, and websites. In addition, entities that are typically more difficult for machines to interpret, such as content and images, were used to complete the network mapping. Using open source intelligence (OSINT), the human trafficking operation was uncovered, comprised of over 350,000 escort advertisements spanning almost a decade. The network operated on three continents, in over fifty cities, and had 30,000 customers. It is concluded that purely OSINT can identify specific individuals and far more criminal activity than previously believed. An operation on the magnitude of the Chinese organized crime network studied could not be successfully identified and indicted without a proactive exploration into the Memex dataset. This case reveals the need for large-scale data analytics to address large cyber-facilitated crime networks.

Keywords: internet, open source intelligence, organized crime, escort advertisements, Backpage, criminal investigations

Investigaciones de la red de trata de personas: el papel de la inteligencia de código abierto y el análisis de datos a gran escala en la investigación del crimen organizado

Resumen

Este documento examina la migración de la trata con fines de explotación sexual a la web y explora técnicas de investigación de código abierto, herramientas analíticas y conjuntos de datos utilizados para descubrir una red china del crimen organizado dedicada a la trata de personas. Memex, un programa de investigación del gobierno de EE. UU. Que produjo un gran conjunto de datos y herramientas de aplicación de software, proporcionó inteligencia web superficial y profunda a través de anuncios de escolta y publicaciones en foros de revisión de compradores sexuales a investigadores de la policía. Las herramientas proporcionaron visualizaciones para explorar las relaciones entre anuncios en línea aparentemente dispares que utilizan atributos, como números de teléfono, direcciones de correo electrónico y sitios web. Además, las entidades que suelen ser más difíciles de interpretar para las máquinas, como el contenido y las imágenes, se utilizaron para completar el mapeo de la red. Utilizando la inteligencia de código abierto (OSINT), se

descubrió la operación de tráfico de personas, compuesta por más de 350,000 anuncios de escolta que abarcan casi una década. La red operaba en tres continentes, en más de cincuenta ciudades y tenía 30,000 clientes. Se concluye que puramente OSINT puede identificar individuos específicos y mucha más actividad criminal de lo que se creía anteriormente. Una operación sobre la magnitud de la red china del crimen organizado estudiada no podría identificarse y procesarse con éxito sin una exploración proactiva del conjunto de datos de Memex. Este caso revela la necesidad de análisis de datos a gran escala para abordar grandes redes delictivas facilitadas por el ciber.

Palabras clave: Internet, inteligencia de código abierto, crimen organizado, anuncios de acompañantes, Backpage, investigaciones criminales

人口非法交易网调查：开源情报与大范围数据分析在调查有组织犯罪中发挥的作用

摘要

本文检验了性剥削非法交易在网络上的流动，并探究了用于揭开一个涉嫌人口交易的中国有组织犯罪网而使用的开源研究技术、分析工具、数据集。Memex，一个生产大型数据集和软件应用工具的美国政府研究项目，通过（分析）应召广告和性买家评论论坛帖子，为执法调查人员提供了表面和深层网络情报。工具提供可视化技术，使用电话号码、电邮地址、网络等性质探究看似不相关的网络广告之间的关系。此外，那些对机器而言一般更难以诠释的实体，例如内容与图片，被用于完成网络映射。通过使用开源情报（OSINT），揭开了非法人口交易操作，其由时间跨度近十年的超过350,000个应召广告组成。该交易网络覆盖三个大陆，超过50个城市，拥有30,000名顾客。结论则是，仅使用OSINT能识别特定人员和远比预期更多的犯罪活动。如果不对Memex数据集进行主动探究，则无法成功识别并起诉所研究的中国有组织犯罪网络的操作。本案例揭示了需要用大范围数据分析应对大型互联网犯罪网络。

关键词：互联网，开源情报，有组织犯罪，应召广告，Backpage，犯罪调查

The availability of the internet has allowed for a dramatic expansion of customer access to the purchase of commercial sex and for exploiters to advertise victims of human trafficking. The internet has allowed criminal organizations and facilitators to reach customers in more direct ways while anonymizing the identity of the perpetrators. The online environment was first used on a mass scale by distributors of child pornography, with one distributor's website alone receiving over one million downloads in the late 1990s in a single year.[2] Yet the scale of the trade in child pornography was subsequently overtaken by the proliferation of advertisements for sexual services, many of these offering sexual services with minors, which had a larger potential market than child pornography. A major US government-funded computer research program, known as Memex, reported identified advertisement sales of about $250 million spent on more than sixty million advertisements for commercial sexual services in a two-year period between 2014 and 2016.[3] The ease of use of online advertisements allowed these criminal businesses to expand with rapidity. A new technical approach was needed, as analyzing the criminal advertisements on such a scale could not be done manually by law enforcement.

Distributors of child pornography were among the first to take full advantage of the anonymity and the reach of the internet through online websites in the mid-to-late 1990s.[4] The provision of child pornography to users for pay or for free exchange occurred at a time when the number of global internet users was a small fraction of the present day, an estimated 150 million users in 1998.[5] The United States Customs Service established a center in Northern Virginia in the 1990s to monitor the distribution of child pornography. Its locale was advantageous as most internet service providers (ISPs) were then located in Northern Virginia, meaning at some point in the process of distribution, the child pornography would very likely transit Northern Virginia. The Customs Service, working closely with the Federal Bureau of Investigation, had jurisdiction as federal statute prohibited the production, distribution, and receipt of child pornography.[6] Having a unique window into the distribution of child pornography, it was possible to determine the recipients of the images and the distinct hubs around the world from which large

2 Interview done in conjunction with Louise I. Shelley, "Crime and Corruption in the Digital Era," *Journal of International Affairs* 51, no. 2 (Spring 1998): 615–617.

3 Larry Greenmeier, "Human Traffickers Caught on Hidden Internet," February 8, 2015, https://www.scientificamerican.com/article/human-traffickers-caught-on-hidden-internet/ and also the accompanying visualization that reveals the international links, "Scientific American Exclusive: DARPA Memex Data Map," retrieved August 7, 2019, https://www.scientificamerican.com/slideshow/scientific-american-exclusive-darpa-memex-data-maps/.

4 Shelley, "Crime and Corruption in the Digital Era," 615–617.

5 Internet World Stats, retrieved August 7, 2019, https://www.internetworldstats.com/emarketing.htm.

6 United States Department of Justice, "Citizens' Guide to U.S. Federal Law on Child Pornography," retrieved August 7, 2019, https://www.justice.gov/criminal-ceos/citizens-guide-us-federal-law-child-pornography.

amounts of child pornography were downloaded. Efforts to collect data on the distribution of child pornography can be identified as the beginning of interpreting large datasets in relationship to human trafficking.

Human trafficking in the form of advertisements for escort and other sexual services followed child pornography into the online environment. The prominence of sexual online marketplaces reached a significant scale in the early 2000s. Prior to the widespread adoption of the internet, advertisements for massage, escort or adult services, and personal ads were scattered in the Yellow Pages, the back of magazines, and other periodicals. By the end of 2003, there were already over seven hundred million users of the internet, a population heavily concentrated in the developed world, where internet access arrived first. Early users of the internet were most often affluent and male,[7] which is the typical demographic of a high-frequency purveyor of commercial sexual services.[8] Naturally, the providers of these services—pimps, criminal organizations, and independent sex workers—turned to this new form of communication to reach and expand their customer base. The scale of online advertisements for commercial sex rose significantly with the popularity of Craigslist beginning in the mid-1990s, which provided the first widely used public marketplace for advertising goods and services online.[9]

The web, in this period, provided a hospitable environment for the expansion of trafficking for sexual exploitation because the first federal legislation recognizing human trafficking in the United States was not passed until the year 2000. Many states were also slow to adopt their own laws. Therefore, human trafficking was not a criminal offense in all jurisdictions, unlike child pornography, for which dissemination was strictly prohibited.[10] Furthermore, the online platforms where sexual advertisements were placed were exempted from responsibility for the user content they hosted under Section 230 of the Communication Decency Act of the Telecommunication Act of 1996. This Act provides immunity from liability for providers of interactive computer services that publish information provided by users. This allowed webhosting services to maintain websites that had large numbers of advertisements featuring persons available for sexual services without liability.

Yet, as these advertisements proliferated, they attracted more law enforcement attention. In 2010, Craigslist shut down its adult services section after mount-

7 Pew Research Center, "Internet/Broadband Fact Sheet," retrieved August 5, 2019, https://www.pewinternet.org/fact-sheet/internet-broadband/.

8 Demand Abolition, "Key Facts About Preventing Trafficking Victimization through 'Demand Reduction,'" 2017.

9 Craigslist.org, "Mission and History," retrieved August 5, 2019, https://www.craigslist.org/about/mission_and_history.

10 For federal jurisdiction, human trafficking must occur across state lines; this could not be established in many trafficking cases.

ing pressure from activists and state attorneys general.[11] This led to advertisers of adult services migrating to Backpage, which then became the predominant player in the marketing of sexual services. The public online marketplace, Backpage, differed from Craigslist in that most of its money was generated specifically through the posting of escort and massage advertisements, and therefore, it more freely advertised its adult services sections.[12] Backpage was shut down by federal authorities in April 2018 shortly before greater controls were placed on internet hosting providers as a result of new federal legislation. In the month before it was shut down, Backpage posted over 133,000 advertisements for sexual services.[13]

The ability to legally place advertisements online for sexual services ended in the spring of 2018 when the Stop Enabling Sex Traffickers Act (SESTA) and Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA) legislation was passed by Congress with overwhelming support. The recently enacted legislation removed the protections for webhosting services, ISPs, and social media sites in regard to the crime of human trafficking. The legislation further provided penalties for someone who "owns, manages, or operates an interactive computer service (or attempts or conspires to do so) to promote or facilitate the prostitution of another person." Those found guilty can face up to ten years in prison and hefty fines.[14]

Within a month of the passage of the FOSTA-SESTA legislation and the censoring of Backpage, advertisements for commercial sex plummeted 82 percent, according to an organization mining escort advertisements. However, after four months, the numbers of advertisements jumped back to 75 percent of their daily volume before Backpage was censored.[15] Some analysts believe many advertisements have merely shifted since 2018 to platforms hosted on servers outside the reach of the United States and not subject to the new legislation.[16]

---

11 Will Saletan, "Pimp Mobile: Craigslist shuts its 'adult' section. Where will the ads go now?" *Slate*, September 7, 2010, https://slate.com/news-and-politics/2010/09/craigslist-shuts-its-adult-section-where-will-sex-ads-go-now.html.

12 California Attorney General, "Attorney General Kamala D. Harris Announces Criminal Charges Against Senior Corporate Officers of Backpage.com for Profiting from Prostitution and Arrest of Carl Ferrer, CEO," October 6, 2016, https://oag.ca.gov/news/press-releases/attorney-general-kamala-d-harris-announces-criminal-charges-against-senior.

13 R. Tarinelli, "Online Sex Ads Rebound, Months After Shutdown of Backpage," *Forensic Magazine*, November 30, 2018, https://www.forensicmag.com/news/2018/11/online-sex-ads-rebound-months-after-shutdown-backpage.

14 *Allow States and Victims to Fight Online Sex Trafficking Act of 2017,* HR 1865, 115th Congress, retrieved August 7, 2019, https://www.congress.gov/bill/115th-congress/house-bill/1865.

15 C. Biederman, "Inside Backpage.com's Vicious Battle with the Feds," *Wired*, June 18, 2019, https://www.wired.com/story/inside-backpage-vicious-battle-feds/.

16 E. Heil and A. Nichols, "Hot Spot Trafficking: A Theoretical Discussion of the Potential Problems Associated with Targeted Policing and the Eradication of Sex Trafficking in the United States," *Contemporary Justice Review* 17, no, 4 (2014): 424.

The Memex Project

The pioneer of the internet, Defense Advanced Projects Research Agency (DARPA), part of the Department of Defense, launched Memex in 2014, a major analytical program focused on developing methods for indexing parts of the internet that were previously unsearchable by exploring trafficking for sexual exploitation online. This multi-year program was motivated by both a national security interest and a desire to develop search engine capabilities to help law enforcement use information on the deep and dark web; information that cannot be found by mainstream, surface web search engines. This program, employing many of the country's top computer scientists, cost nearly $67 million[17] and developed key tools to scrape and locate data on the deep and dark web.

Much of the data found by the DARPA research team on commercial sex advertisements was located on the surface and deep web, as opposed to the dark web, given the desire of commercial sex advertisers to reach the largest number of customers. Customers could easily access advertisements for sexual services on the surface web featured on the specialized sections of Craiglist and Backpage that featured links to more exclusively sexual websites, such as Escortphonelist, escort review sites existing behind paywalls on the deep web, such as TheEroticReview, and over one hundred other identified websites containing erotic, massage, and sexual advertisements. Advertisers of sexual and erotic services generally avoided use of the dark web for several reasons. First, the dark web is not easily accessible by the typical sex buyer. In order to access the dark web, the user must know the specific web location of the site or marketplace they want to enter. Secondly, the user must use a specialized web browser, such as Tor, to access the dark web, which surpasses the computer skills of many possible customers. In general, use of the dark web is in opposition to the desires of those seeking sexual services, who want variety, choice, and the ease and speed of accessibility; therefore, advertisers prefer to use sites on the surface or deep web.

A central objective of the Memex project was to harness open source intelligence (OSINT) from escort advertisement websites and sex buyer review forums by storing the intelligence in a searchable database and equipping investigators with otherwise inaccessible information. This intelligence could more readily inform investigators' use of criminal intelligence contained in their sensitive, closed databases. In order to accomplish this goal, Memex had to provide an efficient means for searching large and diverse types of content that could not be analyzed manually. Additionally, the data collected was frequently changed or manipulated by the poster, in order to avoid detection. The collected and searchable OSINT provided additional opportunities to cross-reference information generated by

17 C. Pellerin, "DARPA Program Helps to Fight Human Trafficking," *DoD News, Defense Media Activity*, January 4, 2017, https://dod.defense.gov/News/Article/Article/1041509/darpa-program-helps-to-fight-human-trafficking/.

leads, to understand the elements of the network, and to provide a larger context to place the activities of the criminals and groups that sold commercial sex.

DARPA transitioned operation of the Memex database in 2018 to an organization funded by the District Attorney of New York, using monies seized from criminals to maintain the database and continue scraping the web. The Memex database, now known as Tellfinder, is composed of daily crawled and scraped escort advertisements and sex buyer review forum posts, featuring over one hundred different sites and sources on the publicly available surface and deep web. These sources are available in multiple languages from international websites. Tellfinder is employed by many law enforcement agencies on the federal, state, and local level.[18]

Reactive vs. Proactive Investigations

Before the advent of Memex and the use of advanced processing technologies, investigations of perpetrators of trafficking for sexual exploitation were limited to those provoked by leads related to an event, report, or tangential evidence connected to another crime. Traditional police work could not efficiently identify the diverse networks of a larger human trafficking enterprise with operations spanning numerous cities. With the new tools of Memex, complex networks, and even supply chains, could be identified, as the subsequent case study illustrates.

Without tools to provide macro-level analysis, criminals have an asymmetric advantage. Criminals can reach large numbers of customers by posting numerous advertisements across platforms, while law enforcement is limited in its capacity to uncover large bodies of advertisements, unless unlocked by data analytics and advanced processing tools. Using these tools to generate tactical intelligence from large data allows law enforcement to document the operations of a criminal perpetrator or organization.

Macro-level analysis of advertisements for sexual services has resulted in an important shift in investigating trafficking for sexual exploitation. Previously, proactive investigations focused on luring potential buyers of commercial sex using "stings," typically baiting potential buyers with an escort advertisement posted and operated by law enforcement officers. The focus of such investigations was on the buyer of commercial sex with the intention of curbing demand. This is an important component of combating trafficking for sexual exploitation when looking holistically at the problem, but has limited effect in curbing the criminal activity given the sheer number of buyers.[19] Proactive investigations are also used by crime analysts who sift through escort advertisement sites on a daily basis looking for

18 E. Hall et al., "TellFinder: Discovering Related Content in Big Data," 2015, https://www.researchgate.net/publication/317411475_TellFinder_Discovering_Related_Content_in_Big_Data.

19 Demand Abolition, "Who Buys Sex?: Understanding and Disrupting Illicit Market Demand," 2018, 4.

runaway/homeless youth, missing persons, and other vulnerable populations at a higher risk for recruitment by traffickers and for sexual exploitation.[20]

Memex, however, opened up a new and important way to access data pertaining to the criminal networks behind larger human trafficking operations, illustrated by the following case study. As demonstrated, OSINT can be used to map the operation of organized crime networks recruiting victims from a specific source country, and the advertisement, movement, and exploitation of those victims in destination countries.

Case Study of the "Supermatchescort" Trafficking Network

A recent indictment of a major criminal network began with the investigation of a single Backpage advertisement in San Francisco, California, advertising an outcall escort agency featuring Asian women of different ethnicities. This advertisement contained two telephone numbers with one designated as a telephone number for customers to text message in order to schedule appointments. This feature was interesting as a possible signifier of a criminal network for two reasons: it suggested a higher level of business volume and organization and the use of multiple telephone numbers provided additional opportunities to establish connections to other entities.

To investigate the network, an initial query was made to the Memex database using the telephone numbers from the first advertisement in San Francisco. This inquiry connected the original advertisement to other Backpage advertisements that subsequently provided more context for the potential illicit operation and additional data points. The network could not be unraveled solely by telephone numbers connections in active advertisements since posters, after advertisements expire and occasionally while they are still active, change the content, contact information, and/or location of the individual or persona advertised before reposting. This is done to both avoid detection by law enforcement and reach new customers. However, it also generates large interconnected webs of entities suitable for analysis.[21]

For example, the Backpage advertisement under investigation illustrates this principle as it was changed by the poster after its original publication to include a new telephone number and a new social media handle. The revised advertisement, identified by Memex's algorithm on extracted data, linked many additional advertisements to the initial advertisement in San Francisco because each connected advertisement contained an entity, such as a telephone number, from another version of the original advertisement through one or more degrees of separation.

20 Heil and Nichols, 423.

21 S. Yu, "Human Trafficking and the Internet," in *Combating Human Trafficking: A Multidisciplinary Approach* (Boca Raton: CRC Press LLC, 2014), 70.

Extensive analysis was required to assemble the network since the advertisements revealed many common patterns that could be ascertained only by iterative analysis and comprehensively combing the escort advertisement dataset for potentially connected data points. The Backpage advertisements of the network, all featuring Asian women, had an easily interchangeable format in the free text portion of the advertisement, featuring descriptive language of different individuals, allowing the poster to replicate them quickly. Beyond the connecting entities, such as telephone numbers, email addresses, and embedded hyperlinks, commonly noted language descriptors of the network, such as unique persona descriptions, language advertising specials, common misspellings, and poor grammar, were used to further construct the network and its operations.[22] Given the limitations of machine learning classifiers in regards to textual and image analysis, these more subjective indicators of the network were often classified manually.

Analysis of both the designated call and text telephone numbers of the linked advertisements through telephone number registration searches revealed that almost all of the entities used Voice over Internet Protocol (VoIPs) providers. Initially discovered telephone numbers were all registered to the same VoIP provider, however, after fully mapping the network, the operators were found to have used additional VoIP providers. Of these numbers, some of the designated text telephone numbers were registered to a VoIP provider specializing in SMS infrastructure for call centers, providing companies and organizations with methods to communicate with customers through text message and other application-based messaging platforms. When mapping the network, the entities with the largest number of connections were two WeChat IDs. WeChat is a popular application-based messaging platform among Chinese populations in China and abroad.[23] From this information, it was inferred that application-based messaging platforms, particularly WeChat, and VoIP services were used to schedule and deploy outcall escort services, potentially from a centralized location. This evidence supported the conclusion that this network was highly organized, used technology to efficiently coordinate activity across multiple cities, and was operated by Chinese individuals, potentially from outside the United States.

Distinct personas of individuals described in the advertisements began to surface through analysis, although these were suspected to be fake personas. Image analysis conducted using Google image search of the photographs featured in the Backpage advertisements and escort agency websites, described below, displayed

22 Elements used to link the advertisements included telephone numbers, designated "text only" telephone numbers, email addresses, WeChat IDs, strings of unique text, the content structure of the advertisement, images and photoshopped features within images, advertisement post ID numbers, locations and physical addresses, persona descriptions, listed external websites, and identities of potential perpetrators.

23 R. Hollander, "WeChat has Hit 1 Billion Monthly Active Users," *Business Insider*, March 6, 2018, https://www.businessinsider.com/wechat-has-hit-1-billion-monthly-active-users-2018-3.

commonalities between the mapped network nodes. First, within the network, a vast majority of personas used professionally taken photographs of lingerie models, which is common in advertisements for illicit commercial-front enterprises. Given the high cost of professional photo shoots, these images are commonly copied from original sources and used to entice potential buyers with attractive women in glamorous settings. Later analysis of the escort agencies' listings on sex buyer review forums confirmed that individuals arriving for outcall appointments did not match the personas featured in the advertisements. This can be considered an indicator of trafficking for sexual exploitation because of the element of deception given the false advertisement and the level of organization in transporting multiple victims to outcall appointments. Second, the image analysis revealed that the advertisements in this network were frequently manipulated with image altering software that obscured the facial features of the pictured individuals. Three methods were used: blurring the eyes or face, covering the eyes or face with emojis, or obscuring the face with a white spot that mimicked a bright light shining on the face. Third, the images were also used as another method for conveying contact information for setting up appointments, featuring superimposed telephone numbers somewhere in the image. The specifics behind the use and manipulation of the images provided additional evidence that the advertisements were connected.

Additional analysis led to the identification of travel patterns of personas, further demonstrating the connection between operations in different cities and countries, as personas would follow similar route patterns. The pattern flowed as follows: Sydney, Australia to San Francisco, California to Vancouver, Canada to Calgary, Canada to Toronto, Canada, then personas dispersed to different cities in the Midwest and eastern portions of the United States.

Many of the Backpage advertisements featured an embedded hyperlink that linked to an external website for local outcall escort agencies solely featuring Asian women. The websites of the assembled network all featured the same basic layout, content, and site structure, suggesting they might be owned and operated by one group or an individual webmaster. Most of the network's Backpage advertisements linked to an escort agency website solely catering to the advertisement's listed city; however, escort agency websites for potential hub cities were sometimes used instead. These repeated escort agency website locations indicated centralized nodes in the mapping of the network, which functioned similarly to the designated text message telephone numbers and two WeChat IDs in fully connecting the reach of the network. A WHOIS domain registration search was conducted on the escort agency websites extracted from the Backpage advertisements. The results established that two individuals owned all of the web domains, in addition to revealing more escort agency sites, totaling fifty-five websites in the United States, Canada, and Australia. A registered physical address in Toronto, Canada was also revealed. Using open web searches, it was discovered that the address was linked to a travel

agency specializing in travel between China, the United States, and Canada. At this time, it is unknown what role, if any, this travel service played in the operation.

Through mapping the entities and other strings of content using OSINT, the identified network was linked to an additional 350,000 escort advertisements and fifty-five websites in more than twenty-five cities in the United States, Canada, and Australia. Eventually, these hundreds of thousands of Backpage escort advertisements were identified with one key individual, a facilitator.

Discussion of the Case

Two years after the submission of the OSINT network analysis stemming from the San Francisco Backpage advertisement to a federal law enforcement agency, seven Chinese nationals were indicted in relation to the network in November 2018. According to the indictments, the "Supermatchescort" network operated in fifty-eight cities throughout the United States, Canada, and Australia, primarily through escort advertisements on Backpage and escort agency websites. In the OSINT network analysis conducted, half of these city-based operations were identified.[24] In addition, with the OSINT provided to law enforcement, the Federal Bureau of Investigation seized more than five hundred web domains associated with the operation of the network.[25]

The indictments confirmed many of the suspicions around the supply chain for the network, its operation, and the trafficking activity found through the OSINT network analysis. The primary defendant supervised a cadre of localized dispatchers stationed in each city and orchestrated the flow of people and money throughout the United States, Canada, and Australia. He facilitated the procurement of property for residential brothels, the recruitment of victims from China through WeChat, and enforced a communication system through WeChat specific to the trafficking operation in each city for locally based dispatchers. Through WeChat, this key facilitator provided the mechanism used by the locally based dispatchers to schedule appointments with sex buyers, transport victims, and facilitate trafficking in sexual exploitation. WeChat was essential to the operation of this criminal enterprise and shows the role of emerging technologies in current human trafficking operations.

The primary defendant implemented a customer relationship management database that contained over thirty thousand records of commercial sex buyers. These records retained customer contact information and details regarding specif-

24 *United States of America v. Zongtao Chen a.k.a. Mark Chen, Weixuan Zhou, Yan Wang a.k.a. Sarah, Ting Fu, Chaodan Wang,* November 15, 2018, https://www.justice.gov/usao-or/press-release/file/1124296/download.

25 Department of Justice, US Attorney's Office, District of Oregon, "Nationwide Sting Operation Targets Illegal Asian Brothels, Six Indicted for Racketeering," January 16, 2019, https://www.justice.gov/usao-or/pr/nationwide-sting-operation-targets-illegal-asian-brothels-six-indicted-racketeering.

ic victims and the services they provided for each scheduled appointment. The indictments further allege the secondary defendant owned the web domains discovered early in the OSINT network analysis. These two defendants operated much of the network remotely, without ever interacting with victims or customers. Further down the operational hierarchy were the dispatchers and on-the-ground operators of the residential brothels, who shuttled victims to appointments, facilitated transportation of victims between cities, and maintained continuous business operational needs.

This case demonstrates that an investigation using purely OSINT can almost entirely map illicit networks of activity and identify specific individuals, which is far more reach than previously believed. In this case, the OSINT network analysis provided investigators with an overarching framework to connect more localized, seemingly separate trafficking operations, known previously to law enforcement in several cities, to a larger organized crime network operating on an international scale. The open source data and analysis enabled federal, state, local, and international law enforcement entities to more effectively gather evidence through traditional, closed sources and undercover operations leading to indictment. An operation of this magnitude could not be identified without proactively and iteratively exploring the existing Memex dataset. Much more analysis still must be conducted to fully understand the complexity of operating this vast illicit network and translate our findings into lessons learned for application to future investigations.

Conclusion

The scale of the "Supermatchescort" case—the large number of advertisements, customers, and the dispersion of victims across three continents—demonstrates the growth of human trafficking networks with the rise of new technology and the deployment of websites and social media to recruit victims and advertise to customers. The large number of identified customers—thirty thousand— and the relative low-cost of reaching these customers through online advertisements and free application-based messaging platforms suggests this enterprise was immensely profitable. The sophistication of criminals, in this case a technologically savvy Chinese network, reveals the possibilities of expanding the human trafficking business to unprecedented levels using innovative methods and cyber strategies.

The analysis of this network with the use of OSINT aggregated by the Memex dataset revealed an integrated business from its central figures in China and Canada to the sale of women in Australia, Canada, and the United States. The potential victims of human trafficking in this network were predominantly Chinese, as were the lead facilitators of the online advertisements and the lower-level dispatchers and operators. The operation functioned across continents with the sale of victims contained within this closed network. This network resembles the

smaller human smuggling and trafficking operations first identified by US law enforcement over two decades ago, before these criminals expanded significantly by exploiting the internet.[26] The supply chain of the operation revealed through this advanced network analysis shows a direct link between sales overseas of commercial sex and the key node of the network in China.

Historically, OSINT has played an important, but secondary, role in investigations; however, OSINT, as this case reveals, has the potential to be a primary method of discovery and research, especially in regards to trafficking for sexual exploitation. The development of Memex's dataset and tools profoundly changed the means by which human trafficking investigations have been undertaken in some jurisdictions. With tools developed through Memex, capable of analyzing and categorizing large quantities of data, OSINT obtained by scraping and analyzing enormous amounts of web content has assumed a key role in the discovery of criminal activity, mapping illicit networks and supply chains, and identifying key facilitators and elusive kingpins, specifically within organized human trafficking operations. The OSINT techniques used in identifying this criminal network may be useful in studying other criminal networks beyond the scope of human trafficking activity, given the migration of many forms of organized crime onto the web.

Despite the enactment of SESTA-FOSTA and the seizure of Backpage on which this network was initially detected, the problem of trafficking for sexual exploitation through online advertisement for sexual services has not ended. Although legislation continues to limit where advertisements are posted, exploiters will continue to migrate to new escort advertisement sites, hide behind sites with paywalls, and move to social media and encrypted communication platforms given the profitability of human trafficking. All of these factors suggest OSINT and large-scale data processing and analytical tools will remain important for future investigations as traffickers, and the technology they use, evolve. To continue to conduct these proactive, network-based investigations, technology companies, law enforcement, and government entities, such as DARPA, must continue to invest in new technology to counter criminal networks and provide intelligence to law enforcement. Technology companies, law enforcement, legislators and policymakers need to be prepared for the constant shifting of the criminal space in regards to human trafficking and be able to respond to nimble and innovative criminal enterprises.

26 Louise Shelley, *Human Trafficking: A Global Perspective* (Cambridge and New York: Cambridge University Press), 114–118.