Digital Communications Surveillance: A challenge for Rhetoric Studies

Cezar M. Ornatowski and Akshay Pottathil

1. Introduction

Digital Communications Surveillance (DCS) in its broadest sense includes analysis of any type of digital data. It spans a wide range of activities, from data gathering by Google using the communications of Google's Gmail customers to the monitoring of home and workplace computer activity, analysis of public attitudes regarding products, policies, or persons, and on to searches using a "fusion"[1] of data sources (Internet, social media, camera, satellite, drone, cell phone, and other) to achieve temporal and geo-spatial, multi-dimensional "mappings" of the search domain for security or military/intelligence purposes.

Digital Communications Surveillance has grown in tandem with advances in digital communication technologies. David Lyon considers "information societies" to be, *ipso facto*, "surveillance societies"[2] and sees surveillance as a "key feature of modern life", a flip side of the use of electronic technologies in all areas of human activity.[3]

At one end, DCS is part of the burgeoning market in the generation and processing of information and data and a response (albeit one that raises social and ethical issues) to the changing capacities of and challenges presented by communication technology. In a world where almost any kind of information may be available to anyone at a keystroke and where information is a form of capital, control of and access to information become issues. At the same time, where work is performed to a large extent at a computer (thus is not immediately visible to others) or remotely (*i.e.* in telecommuting mode), control and management of work and work time also become issues.

Examples of DCS for purposes of home or office access control and work management include such commercially available products as

---

[1] According to the National Research Council, "fusion" means "the use of computer technology to acquire data from many sources, integrate this data into usable and accessible forms, and interpret the results", quoted in Hsinchun Chen, Edna Reid, Joshua Sinai, Andrew Silke, and Boaz Ganor, eds., *Terrorism informatics: Knowledge management and data mining for Homeland Security* (New York: Springer, 2009): xv.

[2] David Lyon, *Surveillance society: Monitoring everyday life* (Buckingham: Open University Press, 2001): 1.

[3] *Ibid*. 2.

© African Yearbook of Rhetoric 3, 1, 2012, ISSN 2220-2188:
Cezar M. Ornatowski and Akshay Pottathil, "Digital Communications Surveillance", pp. 13 - 22.

Brickhouse Security's *Cellphone Spy* (which monitors cell phone activity), *Stealth iBot Computer Spy* (which records all computer activity onto a remote flash drive), the *Porn Detector iBot* (which detects potential pornographic content through recognition of facial features, flesh tone colours, and body postures), and the *Key Logger* (which captures all keyboard activity).[4] The need for such devices is typically justified in terms of vulnerable or subordinate relationships (child safety or protection of employees) or company and product confidentiality and workplace security.

Such forms of DCS fall under what Roger Clarke called "dataveillance": "systematic monitoring of people's actions or communications through the application of information technology".[5] Lyon sees such "everyday" surveillance as "the outcome of the complex ways in which we structure our political and economic relationships in societies that value mobility, speed, security and consumer freedom",[6] thus as, paradoxically, a correlative — perhaps even a condition — of our freedom and safety.

At the other, more rhetorically interesting end, DCS combines natural language processing, artificial intelligence, computational linguistics, text analysis, and other data gathering and processing technologies (such as geographic modeling and visualization), to allow analysts to understand, track, predict, and even perhaps control attitudes and behaviors on individual, group, or even global scale. In the contemporary economic, political, security, and strategic environment, words, symbols, and "ideas" constitute critical "information" and their tracking and deployment in networked communications has become a burgeoning space for business, intelligence, and research activity. Such forms of DCS (related terms here include "web surveillance", "information mining", "web content mining", or "data mining") provide a potential new space for rhetoric. It is these forms of DCS that will be the main focus of the remainder of this essay. They include sentiment analysis and other techniques of textual analysis (or "rhetoric data mining"— a forward-looking term[7] that includes language-focused approaches to DCS) associated with opinion research, security, and intelligence.

2. Sentiment analysis

Sentiment analysis (or opinion mining) refers to extracting information on subjective states; it aims to "determine the attitude of a speaker or a writer

[4] http://www.brickhousesecurity.com/computer-surveillance-anti-spyware.html: accessed 11 November 2011.
[5] Lyon, *Surveillance society*, 143.
[6] *Ibid*. 2.
[7] Akshay Pottathil, "Understanding rhetoric data mining and predictive analytics for Homeland Security", unpublished manuscript (San Diego State University, 2008).

with respect to some topic or the overall contextual polarity of a document".[8] The attitude may be expressed as a judgment, affective (emotional) state, or intended emotional effect on the audience. Sentiment analysis is fast becoming a major tool for gauging public opinion concerning products and services as expressed, for instance, in social media or blogs. Analysing on-line consumer opinion has become a "red hot" tech trend, "a kind of virtual currency that can make or break a product in the marketplace", according to a *New York Times* article on the topic.[9]

Sentiment analysis involves two operations: first, generating sentiment lexicons (lists of positively or negatively marked words; one may generate separate lexicons appropriate to the different spheres of experience to be analysed: politics, technology, and so on) and, second, using the lexicons to analyse the "sentiments" contained in the text corpus. Lexicons are generated from "seed lists" of key terms by using algorithms to recursively query for synonyms using *WordNet*.[10] Analysis of the corpus involves assigning positive and negative values, such a +1 or –1 (or +2 or –2 in case of strong positives or negatives, such as "very good"), to occurrences of relevant terms from the lexicon and assessing the overall "sentiment" score for the corpus in respect to the "target" entity. The target entities and expressions of "sentiment" are identified by parsing (analyzing a text in terms to a given formal grammar). The final "sentiment" score may be refined through data-cleaning operations (such as elimination of quotations and duplications between texts) and interpolation with additional contextualizing indicators (cultural values or "world-happiness" indicators).[11]

Beyond gauging the range of opinion on some "target" topic or issue, sentiment analysis can gauge the general "mood" or "zeitgeist" of the times (especially by mining Twitter chatter), and may have predictive applications (for instance, how specific news may affect a company's stock price or how a given candidate may fare in an election). With the addition of spatial

---

[8] http://en.wikipedia.org/wiki/Sentiment_analysis : accessed 21 October 2011.

[9] http://www/nytimes.com/2009/08/24/technology/internet/24emotion.html?pagewanted=all : accessed 21 October 2011.

[10] *WordNet* is a public domain on-line lexical database for English, where words are grouped together by semantic relations into synonymous groupings ("synsets"). See http://wordnet.princeton.edu .

[11] Namrata Godbole, Manjunath Srinivasaiah, and Steven Skiena, "Large-scale sentiment analysis for news and blogs", in *Proceedings of the 2007 international conference on weblogs and social media (ICWSM)* (Boulder, CO.), http://www.google.com/search?hl=en&source=hp&q=largescale+sentiment+analysis+for+news+and+blogs&oq=Large-Scale+Sentiment+&aq=0&aqi=g1g-v1&aql=&gs_sm=c&gs_upl=591l7891l0l11817l22l18l0l7l7l0l821l3914l3-4.2.0.2l8l0 : accessed November 12, 2011. See also Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis", in *Foundations and trends in information retrieval* 2, 1-2, (2008): 1-135.

analysis, sentiment analysis can provide "sentiment maps".[12] This is where sentiment analysis differs from opinion polls, which are time-stamped and based on limited samples. Sentiment analysis provides samples in the hundreds of millions or even billions, making up in volume what it lacks in analytic 'quality'.

It is important to emphasize the role of the human analyst and of domain expertise in constructing the initial list of "seed" terms appropriate to the corpus and purpose, in constructing and refining the algorithm that generates the expanded lexicon (for example, "awe-inspiring" may mean either "dazzling" or "frightening", depending both on context and on the speaker and tone — formal, slangy, or sarcastic), and then in refining the lexicon and evaluating the output. The process is recursive and involves cycles of data mining alternating with cycles of "human curation".[13]

The process is a kind of rhetorical analysis (in the case of sentiment analysis a kind of modified cluster analysis) but directed at very large corpori, mediated by technology (not only in the sense of hardware and software but also in the sense of "technique"[14] embodied in the coding and in the evolving algorithm), and focused on specified purposes. However, it is also important to emphasize that, unlike rhetorical analysis, which offers a "finite" description of a text from some theoretical perspective (for example, neo-Aristotelian), sentiment analysis (or opinion data mining) is a continuous process (which is the difference between an analysis of, for instance, a politician's speech, a one-time event, and of the politician's popularity) and includes a range of texts (newspapers, blogs, social media) that would present a challenge, as an aggregate, to more traditional rhetorical analysis.

Given the present state of the art, the tradeoff in such analysis is between accuracy and meaningfulness, or, put differently, between quantitative and qualitative aspects: the more strictly quantitative the search, the more potentially representative (accurate) the resulting data, but potentially lacking in qualitative meaning. On the other hand, qualitative analysis (typically confined to smaller, selected corpora), while rich in potential meaning, lacks the accuracy gained by mining large corpora. To try to bridge the gap between quantity and quality, some researchers suggest including "fuzzy" rules to incorporate expert domain knowledge and qualitative insights and increase the "learning" capacity of the system.[15]

Sentiment analysis has been used primarily for commercial and research purposes; however, it is also finding its way into another major area

---

[12] Godbole *et al.*, "Large-scale sentiment analysis".

[13] *Ibid.* For a discussion of some of these challenges.

[14] We treat "technique" here in the sense elaborated by Jacques Ellul in *The technological society* (New York: Vintage, 1964).

[15] Rudolf Kruse, Detlef Nauck, and Christian Borgelt "Data mining with fuzzy methods: Status and perspectives", PDF available through *Google*.

of DCS: security and intelligence.

3. DCS and security

The attacks of 9/11, and the subsequent rise of what some have called the "security state",[16] with its need to monitor threats to the homeland and to US interests worldwide, gave new importance to DCS. A parallel development was the post-Cold War advent of "netwar"[17]: conflicts in which revolutionaries, terrorists, extremists, international criminal organizations, political or religious movements, and other non-state actors create alliances and ideologies largely maintained and disseminated via the Internet and other communication technologies and often recruit followers and coordinate actions using these technologies.[18] In an influential dissertation, Amir Dia argued that in the networked communication environments and dispersed social (increasingly global) conditions of netwar, conflict management "increasingly involve[s] *information operations* and *perception management*" [19] as "the capacity of any activism to ensure effective performance may depend… on the existence of shared principles, interests, and goals"[20] and on the technical means to communicate them.[21] More recently, the events of the "Arab Spring" underscored the mobilising potential of media such as Facebook or Twitter.

The Internet (especially the part of it referred as the "Dark Web") became the forum for extremist groups for propaganda, relationship building, communication, fundraising, and recruitment, especially with the shift toward the "lone wolf" strategy and toward recruiting native or local converts (after the intensification of travel restrictions). Both the Times Square bombing suspect Faisal Shahzad and the accused Fort Hood shooter Maj. Nidal Hassan have allegedly been inspired by Internet postings.[22] In a June 2010

---

[16] Diana Priest and William M. Arkin, *Top secret America: The rise of the new American security State* (New York: Little, Brown, 2011).

[17] John Arquilla and David Ronfeldt, *The advent of netwar* (Santa Monica, CA: RAND Corp., 1996): monograph MR-789: http://www.rand.org/pubs/monograph_Reports /MR789/ : accessed 15 April 2010. See also John Arquilla and David Ronfeldt, *Networks and netwars: The future of terror, crime and militancy* (Santa Monica, CA: RAND Corp., 2001).

[18] See also Mary Kaldor, *New and old wars: Organised violence in a gobal era* (Stanford, CA: Stanford University Press, 2007).

[19] Amir Dia, *The information age and diplomacy: An emerging strategic vision in world affairs* (Boca Raton, FL: Dissertation.com, 2006): 247, emphasis in the original.

[20] *Ibid*. 250.

[21] See also Michael J. Waller, *Fighting the war of ideas like a real war* (Washington, D.C.: Institute for World Politics, 2007).

[22] "Al Qaeda's media war: From fax to Facebook and Twitter", http://aawsat.com/english/print.asp?artid=id18200 : accessed 26 Sept. 2009.

talk to the American Constitution Society for Law and Policy, Janet Napolitano, US Homeland Security Secretary, argued for Internet monitoring as a tool needed to fight home-grown terrorism threats.[23]

In the context of security, DCS falls under OSINT (Open Source Intelligence).[24] For the purposes of this review, we are focusing on language, text (including video text), and "idea"-focused DCS, not on other types of data mining and/or database cross-referencing, such as, for example, the Computer Assisted Passenger Profiling System (CAPPS).[25]

Systematic approaches to open source DCS typically consists of two steps: identifying and "harvesting" relevant websites or other sources of data and analyzing them. Once relevant websites are identified and/or captured, two types of studies can be applied to them: hyperlink analysis and content analysis.[26] Hyperlink analysis uncovers relationships between communities, either in terms of the strength of relationships (relational analysis) or relative popularity of entities (evaluative analysis) [27] and may focus on communication with known or suspected IP addresses, traffic to and from specific locations,

---

[23] "Napolitano: Internet monitoring needed to fight homegrown terrorism", http://www.foxnews.com/politics/2010/06/18/napolitano-internet-monitoring-needed-fight-homegrown-terrorism/ : accessed 18 October 2011.

[24] Other kinds of intelligence include MASINT (measurement signature intelligence), FININT (Financial Intelligence), HUMINT (human intelligence), TECHINT (technical intelligence), SIGINT (Signal Intelligence), IMINT (Imagery Intelligence), and GEOINT (geo-spatial intelligence).

[25] The 2007 *Data Mining Reporting Act* (under which the director of National Intelligence is obligated to present the US Congress with an annual *Data Mining Report*) defines data mining as "a program involving pattern-based queries, searches or other analyses of one or more electronic databases… to locate a predictive pattern or anomaly indicative of terrorist or criminal activity on the part of any individual or individuals", where the queries are "not subject-based and do not use personal identifiers of a specific individual, or inputs associated with a specific individual or group of individuals, to retrieve information from the database or databases" and where the purpose of such queries, searches, or analyses "is not solely… detection of fraud, waste, or abuse in a Government agency or program" or the "security of a Government computer system". Quoted in the *2010 Data Mining Report*, the latest version as of this writing, covering the period from 1 January to 31 December 2010, pp. 1, available at http://www.au.af.mil/au/awc/awcgate/dni/data_mining_report_for_jan-dec-2010.pdf : accessed 19 October 2011.

[26] A list of sites that track terrorist and extremist websites and that may serve as a starting point for an analysis is provided in Jialun Qin, Yilou Zhou, Edna Reid and Hsinchun Chen, "Studying global extremist organizations' Internet presence using the dark web attribute system", in H. Chen *et al.*, eds., *Terrorism informatics:* 237-266, [240-241].

[27] Hsinchun Chen, Jialun Qin, Edna Reid, Yilu Zhou, and Marc Sageman, "Case study of jihad on the web: A web mining approach", in H. Chen *et al.*, eds., *Terrorism informatics*: 221-235.

patterns of reference to concepts, persons, or events, and so on. Content analysis focuses on key words, word clusters, names, locations, term or reference usage patterns, or other "codable" elements of texts, typically using available search engines such as *Google*, *Yahoo*, or the *Copernic Suite* (comprised of *Copernic Agent Professional*, *Copernic Tracker*, and *Copernic Summarizer)*.[28] A researcher may also design a specialized web crawler of their own.

Chen *et al*, for instance, performed content analysis of a corpus of 39 terrorist websites in terms of six high-level attributes: communications, fundraising, sharing ideology, propaganda for insiders, propaganda for outsiders, and character of the virtual community, with each attribute associated with a set of "low level" attributes (for instance, the "propaganda for insiders" included slogans, dates, martyrs, leaders, banners and seals, and narratives of operations and events as associated low-level attributes).[29] They developed a set of coding schemes to identify the presence of each attribute in a website along with weight scores and visualized the results in "snowflake" diagrams. In addition, by calculating similarity measures between all pairs of websites in the set and developing a scaling algorithm, they visualized and mapped the virtual communities (and their relationships) "hidden" within the set of websites, as well as distinguished between core groups and "sympathisers".

In another study, Qin et al analyzed a corpus of 1.7 million multimedia documents from extremist, terrorist, or criminal organisations and movements using the Dark Web Attribute System (DWAS), which analyzes the appearances of three sets of attributes in the websites — technical sophistication, content richness, and Web interactivity — and assigns each site a score for each attribute.[30] Each set consisted of specific features; for instance, the "content richness" attribute set included the number of hyperlinks, downloadable document, images, audio files, and video files in each website. Qin et al then used the results to compare websites representing different regions and ideologies.

A web surveillance-based, National Science Foundation funded project currently (as of this writing) on going at our university, entitled "Mapping ideas from cyberspace to realspace: Visualizing and understanding the spatiotemporal dynamics of global diffusion of ideas and the semantic web", involves a process typical of advanced DCS: ontology formation, web search, data return and analysis, pattern identification, ontology refinement, and re-

---

[28] Ben E. Benavides, "Targeting tomorrow's terrorist today through open source intelligence (OSInt)", restricted access. See also "Information warfare in urban combat", *International Online Defence Magazine* 1 (2006), http://defense-update.com/features/du-1-06/urban-c4i-3.htm : accessed 19 Oct. 2011.

[29] Chen *et al.*, "Case study of jihad on the web", in H. Chen *et al.*, eds. *Terrorism Informatics*, 221-235.

[30] Qin *et al.*, "Studying global extremist organisations' Internet presence".

peated web search, with parallel mapping of the data onto a spatio-temporal global display that traces emerging foci, paths, and diffusion patterns of information about events (human or natural crises such as epidemics) or "ideas". Ontology-building tools include tagging (coding each word according to part of speech), Named Entity Recognition (which labels names of persons, organisations, or locations), and parsing. The results are mapped (by geo-referencing web addresses, URL, place names, gazetteers, blogs, etc.) over a world map (using GIS tools) with time stamps to provide a visual "information landscape".[31]

DCS proceeds through three progressive stages: data (identification of relevant objects for surveillance and extraction of relevant data), information (placement of data in spatio-temporal contexts), and knowledge (understanding the meaning of the information in terms of specific goals and objectives in order to initiate appropriate action).[32]

A solicitation issued on October 7, 2011 by the US Defence Advanced Research Projects Agency (DARPA) for research proposals focused on "narrative networks" provides an example of deployment of DCS for both research and security. Research funded under the solicitation would involve quantitative analysis of narratives (using, among other tools, web surveillance), understanding the effects of narratives on human psychology and affiliated neurobiology, and modelling, simulating, and sensing of these narrative influences, especially in stand-off modalities, in effect forecasting the potential for "narrative influence" on social actors.[33]

One sub-goal of the project is to "ascertain who is telling stories to whom and for what purpose and to discover latent indictors of the spread and influence of narrative tropes" in social networks, traditional and social media, and conversations in order to "identify the nature of stories" and a "list of necessary and sufficient conditions that… distinguish narrative stimuli from other stimuli", to "identify and explore the kinematics and dynamics of story ontology", including "aspects of narratives that that are universal versus aspects that vary considerably across cultural and social contexts".[34] Another sub-goal calls for identifying the role of stories in influencing political radicalisation and violence, in shaping the process of political negotiation, and in influencing psychiatric or clinical conditions. In order to accomplish this, surveillance tools must decompose narratives to "make them quantitatively analyzable in a rigorous, transparent and repeatable fashion" with a goal of developing narrative analysis tools for studying the

[31] See http://mappingideas.sdsu.edu/ : accessed 26 October 2011.
[32] Edward Walz, *Information warfare: Principles and operations* (Boston: Artech House Computer Science Library, 1998).
[33] See https://www.fbo.gov/download/66c/66c704debb0114a6d1bc03c45c80acbd/DARPA-BAA-12-03.pdf : accessed 26 October 2011.
[34] *Ibid*. 6.

"psychological and neurobiological impact of stories on people" as well as of exploring "how stories propagate in a system so as to influence behaviour".[35]

Information obtained through such DCS will become the basis for research in Technical Area Two, Narrative Neurobiology, whose goal is to understand how stories impact neurobiological processes, from basic neurochemistry to the system and "system-of-systems" levels. Finally, research in Technical Area Three is to develop models and simulations to directly discover, track, and measure "narrative impacts" and predict responses, with the ultimate goal of "prevention of negative behavioural outcomes" and "generation of positive behavioural outcomes".[36] In this last area, the solicitation "strongly encourages" development of "stand-off/non-invasion/non-detectable sensors".[37]

The DARPA project represents a new step in surveillance: from what Lyon sees as the general "disembodying of the persons" into sets of data in "dataveillance" systems to their "re-embodiment" (in terms of knowledge and control of the body) as a result of, or perhaps as a function of, information obtained at least in part through surveillance.

A separate area of deployment of DCS is Information Operations (IO), an aspect of Information Warfare. As part of IO, DCS is used to monitor threats, including potential threats to information systems, as well as to penetrate threat organisations by the insertion of software agents to acquire knowledge of intent, capabilities, and plans. DCS is also deployed to monitor the effects of Psychological Operations (PsyOps) activities and to refine both the message and the delivery media.[38]

4. Conclusion: DCS and rhetoric

Digital Communications Surveillance, in which we include web surveillance, is a "hot" area of research, thanks to its commercial, political, and security applications. While the word "surveillance" awakens associations of "Big Brother" watching, in its broadest sense DCS defines a new arena of data gathering and knowledge production defined by the existence and use of new electronic communication technologies. Lyon suggests that "[t]echnology and society are bound together in a mutual process of co-construction" (he uses the term "technosocial" to express their mutual integration).[39] Especially in the domains of security and military action, but increasingly also in commerce and politics, knowledge gained through DCS influences the

[35] *Ibid*. 7.
[36] *Ibid*. 9.
[37] *Ibid*. 10.
[38] Waltz, *Information warfare*.
[39] Lyon, *Surveillance society*, 23.

decisions and behaviours of actors. Insofar as behaviours (including discursive acts) in turn become subject to surveillance and thus a source of data on which subsequent decisions are based, surveillance becomes a component in the overall "ecology" of decision-making and action, including rhetorical action.

In some of its manifestations, DCS may be regarded perhaps as an extension of what Foucault has called "governmentality" or "biopower", a kind of paternal, bureaucratic, semi-visible but pervasive and largely preventive control characteristic of democratic welfare mass-societies (although in specific "local" instances this control may not be so paternal, resulting, for instance, in drone strikes at "subjects" constituted as hostile by surveillance analysis).

Yet, rhetoric as an analytic perspective and rhetoric scholars as "domain" experts have, to our knowledge, so far had little engagement with DCS. A large part of the reason lies perhaps in the ethical issues surveillance raises, as well as in the relative insulation of most rhetoric scholars from the domains of commerce, security, and especially warfare. Yet, DCS may present an opportunity to engage with large corpori of texts, especially ones that have typically been the focus of rhetoric studies (websites, blogs, social media), with the rhetorical (and not predominantly linguistic) aspects of such corpora, as well as with new categories of rhetorical actors (governments, publics, groups, movements, organisations, and networks) and acts, as well as new domains of discursive activity.

San Diego State University