Section 1

Understanding the Operational Environment

This section looks at how communications technology is changing the environments in which humanitarians work, particularly high-risk contexts. All these articles were selected because together, they raise questions that should be considered when assessing the context in which humanitarian organisations operate.

From a security risk management perspective, context analysis should question how communications technology is changing social dynamics in high-risk environments, as well as how organisations (and others, including hostile actors) gather and distribute information. Are humanitarian organisations being targeted for the data they hold, and how? Which information do we trust and how do we react to an overload of unverified information? If security managers keep referring back to the same sources, are we too suffering from homophily as a sector, and what are the risks this implies? Should our actor mapping also include digital actors, and how do we do this?

Cyber-Warfare and Humanitarian Space

Daniel Gilman

Introduction[2]

Recent publications, notably 'Humanitarianism in the Network Age' from the United Nations Office for the Coordination of Humanitarian Affairs and the 2013 Red Cross World Disaster Report on 'Technology and the Future of Humanitarian Action', have outlined the changing environments for humanitarian work and the potential to use advanced communication systems, 'big data' analytics and other information and communication technologies (ICTs) to transform the way humanitarian action occurs (see Section 2 – Communications Technology and its Impact on Humanitarian Programmes, pp. 32-44). These ideas, from online volunteers providing remote information management support through platforms like the Digital Humanitarian Network,[3] the increased use of biometrics in refugee camps,[4] or real-time tracking systems for cold-chain vaccines, are increasingly a reality.[5]

While offering the potential to improve the efficiency of a response to a crisis, these systems also create new vulnerabilities and ethical and legal challenges, particularly around how to respect and manage privacy. At the same time, many of the same techniques and systems are being increasingly used and co-opted by parties to conflicts, leading to an increase in 'cyber-warfare', politically motivated hacking to conduct sabotage and gather intelligence.[6] While originally cyber-warfare was largely the province of technologically sophisticated countries, like the United States and China, the spread of cheap and easy-to-use technology has fundamentally changed the dynamic in recent conflicts (see Byrne, a. p. 13).

Surveillance and cyberwarfare capacities are now found in many authoritarian regimes, particularly those that also host an international humanitarian presence, notably Ethiopia[7] and Sudan.[8] In addition, recent conflicts have seen the increasing 'para-militarisation' of cyber-warfare, with 'private citizens forming into on-line militia groups to perform cyber-attacks against political opponents'.[9] Evidence from the recent conflicts in Libya,[10] Syria[11] and elsewhere suggests that many of these groups are often linked to governments, but in ways that provide deniability and limit accountability. In some cases, these groups may share overlapping membership with armed groups, wielding guns one day and a laptop the next.[12]

This changing nature of cyber-warfare, particularly as seen in the ongoing conflict in Syria, poses some specific challenges for humanitarians and may largely shape the type of technology and programming that can be effectively used in conflict settings. They also pose a unique challenge to the concept of humanitarian space, understood as the idea that humanitarians can avoid being targeted by belligerents due to their adherence to neutrality and other humanitarian principles.

Increasing vulnerability of humanitarian organisations

The rapid spread of advanced ICTs in humanitarian response has made humanitarian organisations a potential target for different types of cyber-attacks. Humanitarian organisations increasingly store, or are given privileged access to, large quantities of data

2 The views expressed in this publication are those of the author alone, and do not reflect the position of the United Nations. Material for this report was drawn in part from the forthcoming OCHA publication 'Humanitarianism in the Age of Cyberwarfare'. The author would also like to thank John Scott-Railton of Citizen Lab for his expertise and support, without which this paper would not have been possible.
3 See http://digitalhumanitarians.com. [Accessed 1 Sept. 2014].
4 See Kanere. (2013). Classified Fingerprinting. 30 Nov. Available from: http://kanere.org/2013/11/30/classified-fingerprinting; UNHCR. (2012). Modern technology helps meet the needs of refugees in South Sudan. 27 Dec. Available from: http://www.unhcr.org/50dc5a309.html; UNHCR. (2014). UNHCR pilots new biometrics system in Malawi refugee camp. 22 Jan. Available from: http://www.unhcr.ie/news/irish-story/unhcr-pilots-new-biometrics-system-in-malawi-refugee-camp. [All accessed 1 Sept. 2014].
5 UNICEF. (2013). Searching for creative solutions in humanitarian action. 21 Oct. Available from: http://www.unicef.org/emergencies/index_70706.html. [Accessed 1 Sept. 2014].
6 See http://en.wikipedia.org/wiki/Cyberwarfare. [Accessed 1 Sept. 2014].
7 Marczak, B. *et al.* (2014). Hacking Team and the Targeting of Ethiopian Journalists. *Citizen Lab.* 12 Feb. Available from: https://citizenlab.org/2014/02/hacking-team-targeting-ethiopian-journalists. [Accessed 1 Sept. 2014].
8 Marczak, B. *et al.* (2014). Mapping Hacking Team's "Untraceable" Spyware. *Citizen Lab.* 17 Feb. Available from: https://citizenlab.org/2014/02/mapping-hacking-teams-untraceable-spyware. [Accessed 1 Sept. 2014].
9 Ottis, R. (2010). From Pitch Forks to Laptops: Volunteers in Cyber Conflicts. In Czosseck, C. and Podins, K. (eds). *Conference on Cyber Conflict Proceedings 2010.* Tallinn: CCD COE Publications. pp. 97-109. Available from: http://www.ccdcoe.org/publications/2010proceedings/Ottis%20-%20From%20Pitchforks%20to%20Laptops%20Volunteers%20in%20Cyber%20Conflicts.pdf. [Accessed 1 Sept. 2014].
10 Scott-Railton, J. (2013). Revolutionary Risks: Cyber Technology and Threats in the 2011 Libyan Revolution. *CIWAG Case Study Series.* Newport, RI: US Naval War College. Available from: https://www.usnwc.edu/getattachment/01e787b8-ee4c-4efb-8c5a-fe02aa2781ba/Scott-Railton-final-for-website.pdf. [Accessed 1 Sept. 2014].
11 Marczak, W. R., Scott-Railton, J., Marquis-Boire, M. and Paxson, V. When Governments Hack Opponents: A Look at Actors and Technology. *Citizen Lab.* (Unreleased draft).
12 Scott-Railton, J. (2014). Presentation at the 2014 Working Group on Emergency Telecommunications. Available from: http://wget2014.wordpress.com/tag/the-citizen-lab. [Accessed 1 Sept. 2014].

and communications, including phone numbers (for SMS applications), financial information (for cash transfers), fingerprints, iris scans, information on staff and local partners and other information. Humanitarians are also using more two-way communication systems, particularly with SMS and web-based tools like Twitter, to share early-warning and program information and collect feedback. Much of this information is potentially valuable – both commercially and to military or political actors. Humanitarian organisations, many of which have limited ICT expertise to begin with, are often lagging in developing appropriate security protocols. Nor do most organisations conduct privacy impact assessments or use other tools to evaluate the potential risks posed by the data they collect (see Kaiser and Fielding, p. 38).

Beyond criminal activity or fraud, there are a range of motivations to target humanitarian actors: political attacks against the organisations themselves (and what they are perceived to represent, i.e. 'western interests'); to facilitate attacks on communities or ethnic groups who are receiving aid; or to gain access to partner organisations that have provided information or access to networks. While nuisance attacks and vandalism, such as the Syrian Electronic Army vandalism of Human Rights Watch's website,[13] get a larger share of the press due to their public nature, the greatest risk is around data-theft, manipulation and monitoring. The most common attacks use malware like Remote Access Terminals (RAT), which targets are tricked into installing. These can provide almost total access to the target's computer – accessing data, turning on the webcam and microphone, logging keystrokes to identify passwords, manipulating files, etc. (See Byrne, a. pp. 13-16).

Beyond data-theft and surveillance, there are also other emerging areas of risk. One is social cyber-attacks – where people use social media or other communication systems to spread malicious rumours or incite panic. In Assam, India, in 2011, false social media messages, including doctored photos of violence from other situations, were used to convince people that riots and violence were happening in their neighbourhoods, leading to a mass exodus.[14] Humanitarian communications systems, which are presumably highly trusted for their neutrality and relay messages related to disaster and violence, are obvious targets. Hypothetically, a system could be hijacked to send out a warning of an impending attack or disaster, causing displacement without the direct use of force; or military groups could use false notifications of aid disbursements to gather civilians in one place for a terror attack.

Another emerging risk area is attacks on infrastructure systems and devices controlled by computers – the 'internet of things'. Objects with internet connections are recognised as being particularly vulnerable to cyber-attack due to the difficulty in upgrading software and a lack of attention to vulnerabilities until recently.[15] This could pose some unique problems for humanitarian systems. For example, 'smart boxes' that track temperature and location to maintain cold-chain vaccines could be prone to manipulation – resulting in ineffective vaccines being unknowingly delivered. Autonomous unmanned vehicles or delivery systems,[16] life towers that produce flood alerts, or smart toilets[17] that control sterilisation functions are all innovations that are being developed that could be prone to cyber-attacks with potentially serious consequences.

The paramilitarisation of cyber-warfare: the case of Syria

Humanitarian organisations are thus clearly vulnerable to cyber-attacks, and there are benefits for armed groups to consider targeting them explicitly. The role of the Syrian Electronic Army and other groups in the Syria conflict is illustrative of the nature of the way these groups are developing in contemporary warfare.

First, while the hijacking of Twitter accounts and other public advocacy attacks have garnered much of the attention, there is well-documented evidence of systematic attacks on the Syrian opposition and civil society, as well as NGOs. While the tools that are being used are not particularly sophisticated or expensive (the DarkComet RAT that was widely used in Syria to target opposition groups was available for free[18]), a common theme among the attacks has been 'sophisticated social engineering that is grounded in an awareness of the needs, interests, and weaknesses of the opposition.'[19] So, for example, malware has been embedded in tools to protect

13 Fisher, M. (2013). Syria's Pro-Assad Hackers Infiltrate Human Rights Watch Web Site and Twitter Feed. *The Washington Post*. 17 March. Available from: http://www.washingtonpost.com/blogs/worldviews/wp/2013/03/17/syrias-pro-assad-hackers-infiltrate-human-rights-watch-web-site-and-twitter-feed. [Accessed 1 Sept. 2014].
14 Goolsby, R. (Undated). On cybersecurity, crowdsourcing, and social cyber-attack. *Policy Memo Series*. 1. Washington, DC: The Wilson Center. Available from: http://www.wilsoncenter.org/sites/default/files/127219170-On-Cybersecurity-Crowdsourcing-Cyber-Attack-Commons-Lab-Policy-Memo-Series-Vol-1.pdf. [Accessed 1 Sept. 2014].
15 Eisen, M. (2014). The Internet of Things Is Wildly Insecure – And Often Unpatchable. *WIRED*. 4 Jan. Available from: http://www.wired.com/2014/01/theres-no-good-way-to-patch-the-internet-of-things-and-thats-a-huge-problem. [Accessed 1 Sept. 2014].
16 See http://www.matternet.us. [Accessed 1 Sept. 2014].
17 UNESCO. (2014). Smart eSOS toilet for emergencies. 8 July. Available from: http://www.unesco-ihe.org/news/smart-esos-toilet-emergencies. [Accessed 1 Sept. 2014].
18 McMillan, R. (2012). How the Boy Next Door Accidentally Built a Syrian Spy Tool. *WIRED*. 11 July. Available from: http://www.wired.com/2012/07/dark-comet-syrian-spy-tool. [Accessed 1 Sept. 2014].
19 Scott-Railton, J. and Marquis-Boire. M. (2013). A Call to Harm: New Malware Attacks Target the Syrian Opposition. *Citizen Lab*. 21 June. Available from: https://citizenlab.org/2013/06/a-call-to-harm. [Accessed 1 Sept. 2014].

privacy such as Skype encryption or proxy tools, preying precisely on anxieties around cyber-security. Other attacks have included distributing malware through existing social networks, such as hijacking the Facebook page of the 'Revolution Youth Coalition on the Syrian Coast' and posting a malicious link disguised as an investigation of the death of a well-known opposition commander. In other cases, material is designed to be of interest to NGOs or other activists who have connection to opposition groups, such as an NGO administrator receiving an email purporting to contain video evidence of the Syrian military abuses that contained embedded malware. The attacks often come from trusted sources or target private accounts, suggesting 'some degree of prior penetration of the opposition – either through computer network intrusion or other intelligence gathering activities.'[20]

Establishing what direct harm has been caused by these attacks is difficult, but there is circumstantial evidence linking arrests and disappearances to security breaches. Individuals have reported that they were confronted with material from their computers during interrogations, and detainees' accounts are known to have begun seeding malware shortly after their arrests by government forces.[21] All of this suggests that the Syrian cyber-groups are coordinating with military and security services, rather than as *ad hoc* or opportunistic attackers. Like paramilitaries in other conflicts, they have not shown particular respect for international humanitarian law, or for the neutrality of humanitarian actors. For example, there are reliable reports of people communicating with humanitarian organisations over Skype being tortured to give up their passwords, with their accounts then used to transmit malware to NGO staff and their contact networks[22] (see Byrne, a. p.15).

This matters for the way that humanitarians think of these attacks, and how to use information systems. *Ad hoc* attacks or vandalism by 'lone-wolf' hackers may be unavoidable, but will generally pose only limited risk, as these actors are unlikely to be able to act on the information obtained. Systematic targeting of humanitarian information systems and the people who use them by groups linked to military and security actors pose a direct challenge to humanitarian space, however. In particular, given that remote sensing and advanced information networks have been proposed as a way to mitigate access concerns due to increasing attacks on aid workers, the spread of paramilitary cyber-groups should be a worrying development.

The limits of a cyber-security risk management: acceptance in humanitarian cyber-space

Recent surveys and discussions with practitioners suggest that humanitarian organisations have a long way to go to ensure a sufficient level of technical security against cyber-attacks. Most staff are not aware of the nature of the threats faced by field operations, or of basic data security practices, such as how to identify malware attacks (see Byrne, a. pp. 13-16; and Byrne, b. pp. 56-58). There is relatively little use of more sophisticated encryption or security tools; and few if any organisations are working with cyber-security experts to conduct stress tests or monitor for breaches. Precautions like these will probably be a minimum requirement for humanitarians to function in cyber-insecure environments in the near future.

Of course, just as with the use of physical security protection – armoured cars, flak jackets or security guards – the use of a heavily securitised approach can have a negative impact on the acceptance of humanitarian workers, and reduce information sharing and transparency. In the future, humanitarian organisations will need to conduct cyber-security risk assessments to test the basic security of information systems being set up, and also to ensure that there is awareness of the type of threat from cyber-groups. Critically, the level of physical security and cyber-security may not be identical. So a conflict may be relatively secure for humanitarian workers physically, but information systems may be extremely vulnerable (see Byrne, a. pp. 12-16).

If the information coming out of Syria is any model, however, there will be fundamental limits to what technical investments in cyber-security can accomplish. This is because the sophistication in the attacks derives largely from 'social engineering', manipulating people into giving access to their computers, rather than circumventing encryption or other safeguards. Promoting awareness of the nature of threats and regular monitoring of systems can mitigate the risks, as can shifting more work offline or into closed systems. These approaches have obvious limitations, however.

20 *Ibid.*
21 Marczak *et al.* When Governments Hack Opponents. See n. 11 above.
22 Scott-Railton, J. (2014). Digital Security and Wired Humanitarians: Three Trends that Should Scare You. Presentation at the 2014 Working Group on Emergency Telecommunications. Available from: http://wget2014.wordpress.com/tag/the-citizen-lab. [Accessed 1 Sept. 2014].

Instead, it may be more useful to focus less on the nature of the attacks, and more on that of the attackers. With more organised entities, particularly those linked to armed groups, it may be possible to engage in the equivalent of access negotiations to get commitments not to target humanitarian information systems. More broadly, the concept of 'humanitarian cyber-space' could be promoted through negotiations with these groups, online communities, and enlisting local 'white hat' hackers or other online activists. This would require more outreach in local languages, to message boards and online communities, and a more nuanced understanding of dynamics both locally and within the wider diaspora community that may be involved. Of course this would still require regular security assessments to ensure that agreements were being respected, and the difficulty of attributing attacks will make enforcement difficult. Nonetheless, promoting the idea of the neutrality and sanctity of humanitarian information systems may be as effective as any of the other approaches available.

There is also a need for further advocacy on when cyber-attacks on humanitarian organisations constitute a violation of international humanitarian law (IHL). The International Committee of the Red Cross (ICRC) has recognised that cyber-warfare techniques are subject to IHL[23] and Rule 86 of the 'Tallinn Manual on the International Law Applicable to Cyber Warfare', a non-binding study, is that 'cyber operations shall not be designed or conducted to interfere unduly with impartial efforts to provide humanitarian assistance'.[24] Other customary international humanitarian law recognises that objects used for humanitarian relief operations need to be protected from destruction, misappropriation or looting.[25] A case can therefore be made that cyber-attacks on humanitarian information systems, even if only for data-theft, would constitute a violation of IHL as they undermine the ability of humanitarian organisations to deliver impartial assistance. A clearer agreement on what activities constitute a violation of IHL could provide some leverage on governments and non-State actors who might otherwise consider these types of attacks as acceptable, particularly since it is so hard to prove attribution to any specific incident. On this basis, humanitarian organisations should also insist that governments or other belligerents take steps to ensure the cyber-security of activities happening in their area of control.

Conclusions

Humanitarian organisations face a fundamental challenge when considering how to adapt to 21st century conflicts. On the one hand, using the most advanced information systems will allow them to better assess needs, target aid and increase the efficiency of delivery. But the more comprehensive these systems become, the more tempting they become as targets for military and criminal actors. More investments in better cyber-security training, technology and standards are clearly needed to ensure a basic level of robustness in the face of these threats (see Byrne, b. pp. 56-58). However, a highly securitised approach to information systems will be expensive and limit information sharing.

In any case, even the best-designed system will likely be vulnerable to persistent attacks by organised groups, particularly those with strong local networks able to use social engineering or direct coercion. To the extent that cyber-groups are organised or linked to formal armed groups, it is worth considering how humanitarians can engage with them and ensure that the concept of humanitarian space and the neutrality of humanitarian actors is extended to information systems. At the same time, there is an emerging need for more thinking and advocacy to clearly define what constitutes a violation of international humanitarian law in regards to cyber-attacks.

Mitigation and advocacy will only go so far, however. In the end, humanitarian organisations operating in cyber-insecure environments will need to weigh the benefits of setting up certain kinds of information systems, against the possibility that they will be abused or co-opted by parties to the conflict.

23 Furthermore, cyber-warfare does not have to produce permanent, physical destruction to be considered an 'attack'. See ICRC. (2011). *International Humanitarian Law and the challenges of contemporary armed conflicts.* pp. 36-38. Available from: http://www.icrc.org/eng/assets/files/red-cross-crescent-movement/31st-international-conference/31-int-conference-ihl-challenges-report-11-5-1-2-en.pdf. [Accessed 1 Sept. 2014]. See also ICRC. (2013). What limits does the law of war impose on cyber-attacks? 28 June. Available from: http://www.icrc.org/eng/resources/documents/faq/130628-cyber-warfare-q-and-a-eng.htm. [Accessed 1 Sept. 2014].
24 Schmitt, M. N. (ed.). (2013). *Tallinn Manual on the International Law Applicable to Cyber Warfare.* New York, NY: Cambridge University Press.
25 ICRC. (Undated). Customary IHL. Rule 32. Humanitarian Relief Objects. Available from: http://www.icrc.org/customary-ihl/eng/docs/v1_rul_rule32. [Accessed 1 Sept. 2014].

Trends in Intelligence Gathering by Governments

Rory Byrne

Introduction[26]

Advances in digital communication offer many advantages for organisations that seek to do good, such as speed and increased productivity, but also create many new risks such as intercepted communications and systems failure. Humanitarian aid agencies are not immune to either of these effects. While physical security threats and mitigation measures often differ between the human rights and humanitarian sectors, especially with regard to the implementation of security strategies such as acceptance, deterrence and protection, there is a possibility for digital security lessons to be shared – particularly as the humanitarian sector is rapidly increasing its use of technology.

With such a complex topic and such limited space, this article aims to give the non-technical reader an introduction to trends in digital intelligence gathering by governments – though the arguments put forward in this paper equally apply to the use of surveillance and intelligence gathering by non-state actors and private entities.

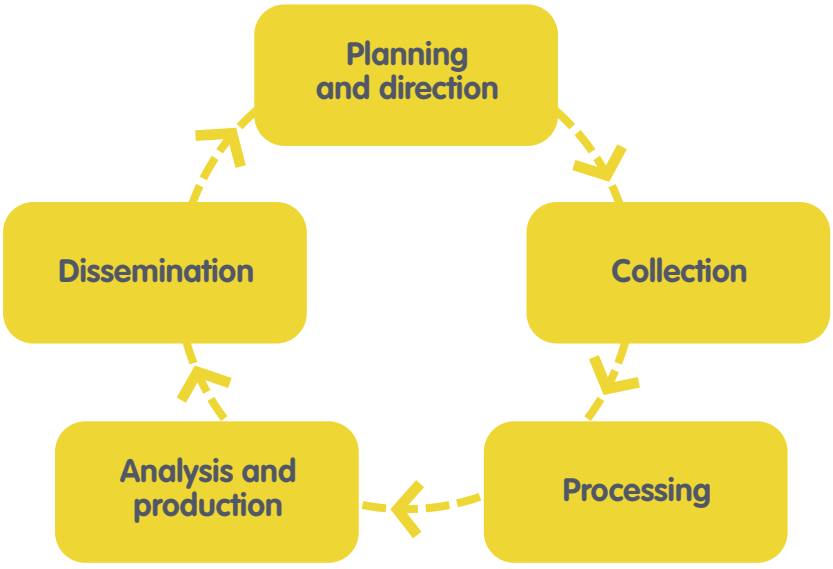
The intelligence cycle

To understand recent trends in digital intelligence gathering by governments, we will utilise the framework of a widely recognised standard to explain how information is gathered and used, overtly and covertly: the 'Intelligence Cycle'.

Planning and direction

Intelligence gathering activities at the governmental level generally begin with requirements set by policy-makers. While it can be argued that some governments, particularly repressive ones, were slow to recognise the threat from – and possible information gathering capabilities of – digital intelligence, developments since the Arab Spring indicate that government planning and direction for digital intelligence is now a common occurrence (see Gilman, pp. 8-10).

Efforts appear to be particularly concentrated around contentious issues such as the emergence of separatists; national groups seeking a change in the balance of power; and/or ad hoc protest movements,



26 The author wishes to thank Eric S. Johnson, Holly Kilroy and a number of anonymous people who graciously agreed to review the article before submission. Any errors or omissions are the author's only.

spurred on by social media and the wisdom of crowds. Security is tightened during critical time periods such as the scheduling/postponing of elections, visits of foreign dignitaries and trade delegations, or civil unrest in a neighbouring regime. The demise of a leader or the fall of a government can lead to a loss of civil liberties – with humanitarian agencies and human rights groups often considered threats that need to be monitored using advanced intelligence gathering methods. Such capabilities are not limited to the larger industrialised powers. Smaller countries such as Belarus, Sudan, Swaziland, Syria, U.A.E. and Vietnam have all been exposed by whistle-blowers and mainstream media as conducting digital intelligence efforts, often thanks, in part, to technical expertise and equipment they receive from governments and corporations.[27]

The increasing prevalence of 'hackers for hire' and the willingness of telecommunications companies to sell communications interception and cyber penetration tools to anyone – regardless of intent – has widely increased the availability of tools, methods and training that can be used not only to attack civilians and non-combatants but also to deliberately and intentionally disrupt the free flow of information by controlling and censoring the internet. Efforts to regulate the export of classified and highly sensitive technologies, by the United Kingdom, the European Union and the United States have been limited due to a range of factors: financial self-interest, dual-use arguments and the desire to 'backdoor' such products for intelligence gathering on the part of the very same countries advocating (publicly) for/against the sale of said products in the first place.

It appears that some organisations are singled out because of the human rights activities they carry out as part of their mandate (e.g. exposing secret prisons), while others are subject to increased scrutiny because of the value of the information they gather (e.g. medical records) (see Gilman, pp. 8-9). For example, Médecins du Monde, together with Amnesty International, UNICEF and WHO, have been targeted by both the Chinese Government[28] and the UK Government Communications Headquarters GCHQ).[29]

Collection

Collection is defined as 'the gathering of raw information based on requirements'.[30] It is the area most commonly focused on in media and other forums, both because of the mystery of 'spying' methods and tools, and because this stage is often the most vulnerable to being revealed, since evidence can often be collected using detection and forensic processes. The focus is often on covert communications intelligence (COMINT); although open source intelligence (OSINT), based on information freely available online, is said to make up the vast majority of final intelligence reports. This is because the raw material is relatively easy to obtain (voluntarily given), highly accurate (based on first person accounts), and rapidly growing in volume and magnitude (connecting the dots has never been easier).

Ironically, the very same tools and techniques associated with open source intelligence gathering represent an important resource for NGOs to help improve their own physical and digital security mitigation measures (see Byrne, b. pp. 56-58).

Background

The first widely publicised incidence of digital intelligence collection against human rights groups was in 2008 (though it is now considered that the alleged abuse(s) may have been ongoing for up to a decade before this) and were linked to Chinese government attacks on Tibetan organisations. This used a method called 'spear-phishing,' a process which involved Chinese intelligence operatives sending fake emails that often appeared to be from internal co-workers (a process known as 'social engineering') and tricked users into opening seemingly innocent documents – which then installed 'trojans' capable of recording all user activity and sending the illegally/illicitly garnered information back to external servers. By targeting the weakest, most vulnerable links – human beings – Chinese intelligence was then able to commandeer an organisation's internal network and establish a long-term capability to monitor all of their public and private communications (known as 'Advanced Persistent Threat').[31] This method continues to be one of the most simple, yet effective, ways of gathering digital intelligence.

27 For an ongoing collection of examples and excellent forensic reports about tools used against activists, see Citizen Lab at the Munk School of Global Affairs, University of Toronto, https://citizenlab.org.
28 Sterling, B. (2012). Amnesty International infested with Chinese Ghost RAT. *WIRED*. 20 May. Available from: http://www.wired.com/2012/05/amnesty-international-infested-with-chinese-ghost-rat. [Accessed 1 Sept. 2014].
29 Taylor, M. and Hopkins, N. (2013). Amnesty to take legal action against UK security services. *The Guardian*. 9 Dec. Available from: http://www.theguardian.com/world/2013/dec/09/amnesty-international-legal-action-uk-security-services. [Accessed 1 Sept. 2014].
30 FBI. Intelligence Cycle. Available from: http://www.fbi.gov/about-us/intelligence/intelligence-cycle. [Accessed 1 Sept. 2014].
31 Kaiman, J. (2013). Hack Tibet. *Foreign Policy*. 4 Dec. Available from: http://www.foreignpolicy.com/articles/2013/12/04/hack_tibet_china_cyberwar. [Accessed 1 Sept. 2014].

Waterholes

A similar vulnerability has been created through the increasing use of a technique referred to as website 'waterholes'. This type of attack works by identifying a website that intelligence targets are known to frequent (for example, a trusted NGO forum) and hacking the website in order to implant malicious pieces of code. When people visit the site with insufficient security (such as poorly maintained or outdated browsers and operating systems) the code can inject 'trojans' onto the user's machine.

Certificates

Another unsettling trend involves the manipulation of the basis upon which much of the security used online (called Secure Socket Layer) to protect web browsers, email, and important transactions depend. These protocols rely on 'digital certificates' (https:// as opposed to http://). The ability to issue false certificates and/or compromise a trusted source (a 'Certificate Authority') has allowed governments, and/or their agents, to impersonate/intercept the day-to-day activities of average citizens.[32] Most users think they have a secure connection to a wide variety of sites and tools – such as Gmail, Yahoo Mail, Facebook, Twitter, WhatsApp, etc.– when in fact, they often do not, as the connection may have been compromised and their data exposed at a number of points along the way (such as at their Internet Service Provider or their Wifi access point).

Mobile phones

Similarly, the technology used to intercept and locate mobile and satellite phones has become cheap and is readily available as COTS (commercial, off-the-shelf) hardware and software – and is suspected of contributing to the death of some journalists in Syria.[33] Phones can serve as tracking devices (even with location services turned off) with similar degrees of accuracy and unbeknownst to most subscribers, and can even be turned on with the microphone activated to allow remote eavesdropping while in off-mode unless the battery is removed. Practically all phone networks have the ability to intercept user calls. For platforms that offer some extra layers of security (such as Blackberry Enterprise Services), a recent trend has been for governments to threaten to or actually block the use and/or sale of such devices and services until the company provides them with a method of intercepting the encrypted data – for example in India and the UAE.

Even when governments cannot intercept the actual content of messages being sent via email and texting, phones generate a significant amount of 'meta-data' – such as location, servers used, sites connected to, time of day, etc.– which means governments already have a strong idea of with whom, where and how you are communicating, even if they don't know exactly what it is being said. Likewise, data generated through social media sites have become a huge reservoir for content-rich intelligence collected by governments and criminal elements because of 'liking and tagging', all done voluntarily.

Phones also pose a security management problem to organisations that want to reduce their exposure to a myriad of risks that stem from the proliferation of hand-held devices, the amount of data being stored, poor security precautions, frequent losses, and the evolution towards cheaper devices (in particular, Chinese-made products found in emerging markets). Recent examples have discovered that some newly purchased phones contain 'backdoors' – such as pre-installed software or hardware which can be used to gain access and control of the device, without the consent of the owner. Discovery of such threats is difficult, if not impossible for most organisations, though the problem can be reduced by sourcing from reputable manufacturers and monitoring phone activity and data usage. This vulnerability is compounded by the growing trend of employees buying and using their own phones, laptops and tablets for work purposes (instead of being issued them by their technical departments). At a minimum, organisations seeking to mitigate such threats should institute effective 'bring-your-own-device' strategies, which install security software onto personal phones to allow organisations to provide a base level of security for the work related information stored on the device (see Byrne, b. pp. 56-58).

Physical access

Collection efforts are not limited to remote digital efforts. Physical access to devices allows unscrupulous operators to take advantage of *ad hoc* situations to gather intelligence data. For example, installing hardware devices such as key-loggers into

32 For example, in Iran. BBC. (2011). Fake DigiNotar web certificate risk to Iranians. 5 Sept. Available from: http://www.bbc.co.uk/news/technology-14789763. [Accessed 1 Sept. 2014].
33 Rayner, G. and Spencer, R. (2012). Syria: Sunday Times journalist Marie Colvin killed in 'targeted attack' by Syrian forces. *The Telegraph*. 22 Feb. Available from: http://www.telegraph.co.uk/news/worldnews/middleeast/syria/9098175/Syria-Sunday-Times-journalist-Marie-Colvin-killed-in-targeted-attack-by-Syrian-forces.html. [Accessed 1 Sept. 2014].

computers or placing covert tracking devices on vehicles. A recent development, seen in many countries, has been the use of covert, and in some instances, overt actions deliberately designed to break into NGO offices and homes, with hardware being taken or destroyed. Many such examples have emerged from places as varied as Belarus,[34] Egypt,[35] Israel,[36] Russia,[37] Vietnam[38] and Zimbabwe.[39] Similarly, persons of interest have found themselves forcibly separated from their devices at checkpoints such as airports, police stations and hotels – where border patrol and law enforcement officers use the opportunity to search, copy, and retrieve information stored on devices. Another recent trend has been for governments (such as Turkey, Uganda, Kenya, UAE) to introduce laws that make digital intelligence gathering easier; for example, requiring that identification must be produced before purchasing a SIM card or instituting laws that force the disclosure of encryption keys. In some countries such as Syria[40] and Sudan,[41] human rights activists have been tortured until they reveal their passwords to social media, email accounts and computers (see Gilman, p. 10).

Processing

With the explosion of digital artefacts created as a result of the continued expansion of the internet, the increased ability of intelligence agencies to process and store large volumes of data indefinitely has been a troubling development. Helped by the decrease in cost of physical storage devices and the increase in sophisticated data-mining software, processing 'big data' (huge sets of data collected and sorted through advanced analysis techniques) has become not only easier, but routine – in fact, the ability to decrypt, recover (even after deletion), translate, tag and measure intelligence for reliability and relevance has increased the ability of analysts to deal with large volumes of data. As such, a trend has emerged in many countries where governments are attempting to 'collect it all'.[42]

Increased processing capability has led to a wider provision – beyond the need to know – of access to intelligence information. For example, in many countries, digital intelligence is no longer restricted to strategic intelligence organisations. Instead, it is now being made available to local law enforcement with the result that this may have changed the nature of interactions between such citizen-based groups and governmental authorities. With digital intelligence becoming increasingly cheap in comparison to large human intelligence (HUMINT) sources and/or physical surveillance operations, a potential exists that lower priority targets like humanitarian NGOs – who are already targeted because they can expose governments – will be subject to increased surveillance and monitoring (see Gilman, pp. 8-10).

Analysis and production

Recent advances in technology have enabled analysts to make use of a wide range of disparate sources. Data collection and processing can be integrated with sophisticated social network analysis tools, which in turn allow junior-level analysts – or any other low level criminal – to compile a fairly intricate picture of the people, locations and organisations a person and/or network interacts with on a daily basis.

Dissemination

How governments have disseminated and used digital intelligence for tactical purposes has not been without repercussions. Particularly prevalent has been the use of disruption instead of direct attacks on individuals and organisations – the theory being that direct attacks create more attention, while disruption can often produce, if not the same result, outcomes that are more manageable. An example would be the use of spurious legal cases to harass and intimidate. Tactically, this often includes the theft of laptops, the confiscation of servers and/or the burning of offices.

Concerned by the lack of predictability associated with open access, many governments have undertaken efforts to block or censor websites and communications devices, temporarily or permanently – for example China, Egypt, Syria and Turkey. During times of unrest, it is not uncommon for governments to try to shut down internet pipelines (as Sudan did in September 2013), thus limiting the free-flow of information. Organisations must prepare for such

34 Human Rights Watch. (2011). *World Report 2011: Belarus*. Available from: http://www.hrw.org/world-report-2011/belarus. [Accessed 1 Sept. 2014].
35 Australian Associated Press. (2013). Egypt NGO says office raided by police. 19 Dec. Available from: http://www.sbs.com.au/news/article/2013/12/19/egypt-ngo-says-office-raided-police. [Accessed 1 Sept. 2014].
36 Ma'an News Agency. (2012). Israeli forces raid NGO offices in Ramallah. 11 Dec. Available from: http://www.maannews.net/eng/ViewDetails.aspx?ID=546800. [Accessed 1 Sept. 2014].
37 Weiland, S. (2013). A Threat to Relations: Germany Irate over Russian NGO Raids. *Der Spiegel*. 26 March. Available from: http://www.spiegel.de/international/europe/russian-authorities-raid-german-foundations-and-ngos-a-890969.html. [Accessed 1 Sept. 2014].
38 BBC. (2012). Vietnamese bloggers deny charges, third in leniency bid. 16 April. Available from: http://www.bbc.co.uk/news/world-asia-17727373. [Accessed 1 Sept. 2014].
39 Karimakwenda, T. (2012). Civil Society Coalitions issue response to police crackdown. *SW Radio Africa*. 8 Nov. Available from: http://www.swradioafrica.com/2012/11/08/civil-society-coalitions-issue-response-to-police-crackdown. [Accessed 1 Sept. 2014].
40 Blomfield, A. (2011). Syria 'tortures activists to access their Facebook pages'. *The Telegraph*. 9 May. Available from: http://www.telegraph.co.uk/news/worldnews/middleeast/syria/8503797/Syria-tortures-activists-to-access-their-Facebook-pages.html. [Accessed 1 Sept. 2014].
41 Author's confidential security debriefing with Sudanese human rights defender subject to the practice.
42 Nakashima, E. and Warrick, J. (2013). For NSA chief, terrorist threat drives passion to 'collect it all'. *The Washington Post*. 14 July. Available from: http://www.washingtonpost.com/world/national-security/for-nsa-chief-terrorist-threat-drives-passion-to-collect-it-all/2013/07/14/3d26ef80-ea49-11e2-a301-ea5a8116d211_story.html. [Accessed 1 Sept. 2014].

eventualities by creating strategies for resilience such as switching to alternative communication channels that can help bypass censorship – for example, Virtual Private Networks (VPNs), Tor (a free software developed for providing increased anonymity and circumvention of restrictions) and the usage of satellite broadband (see Byrne, b. pp. 56-58).

Both democratic and non-democratic governments are using social media to spread propaganda, while also using these technologies to disrupt the activities of groups they perceive to be hostile to them. They accomplish this by spreading discord and false information within and among groups. Examples include collecting data on upcoming events and arresting people during meetings, or publishing propaganda aimed at the groups which creates conflicts and reduces their organisational effectiveness.

Finally, digital intelligence is often disseminated and used for launching human intelligence operations – for example, personal information about web browsing, email and social media activity can be used for manipulation, blackmail and recruiting agents within organisations. Such 'insider threats' continue to play a key role in the intelligence-gathering arsenal deployed by governments. Even more importantly, these techniques are increasingly used not only at local or national offices but are directed towards international headquarters. This author's experience has uncovered that insider threats – like disgruntled employees or paid cover sources like cleaners or security guards – are becoming a common intelligence tactic used against human rights NGOs by governments. Recruitment and management strategies should aim to reduce underlying threat models that undermine trust and create conditions that lead to the evolution of insider threats.

The Dichotomy of Technology in Conflict

Beauty and the Beast

Anahi Ayala

Introduction

In the field of Information and Communication Technology for Development there is often a debate rooted in the dichotomy between the highly enthusiastic view of technology, as an enabler of information exchange that bypasses traditional gatekeepers such as broadcasting media and governmental agencies; and the highly pessimistic view, that focuses on the dangers of technology such as technical gaps, the digital divide and privacy and security threats.

The truth is somewhere in between. Particularly in conflict situations, the reality is much more complicated. On the one hand, technology, and mobile technology in particular, allows for immediate and broad early warning systems to be created in places where real-time communication would previously have been almost impossible (see Porcaro and Walker, pp. 33-36; see also Mayo, pp. 46-50). On the other hand, the way information moves in those contexts can affect the deepening of already existing divisions and the further polarisation of opposing views, where technology enables both an immediacy and increase in volume of material feeding specific viewpoints. One of the most important ways in which these phenomena play out in humanitarian environments today is in the ways in which affected communities use and experience technology, particularly in conflict environments (see Grayman and Anderson, pp. 22-26; see also Sambuli and Awori, pp. 27-31). This article explores the polarising effect of communications systems that are becoming increasingly 'closed'.

From a security management perspective, this same dichotomy is even more accentuated. On one side technology is allowing a broader and larger reach for monitoring security and conversations happening on the ground that can give us real-time insights on risks (see Sambuli and Awori, pp. 27-31); on the other side technology is posing new risks for humanitarian organisations and creating new systems that bypass the usual communication streams and are therefore hidden. The ability to predict violence and provide real-time support in case of violent incidents is strictly related to both our ability to use technology to monitor the situation on the ground, and also to understand how others may be using it to organise violent actions or to create tension.

The link between violence, conflict and technology, especially its use by affected communities and parties to conflict, is only beginning to be understood and the available evidence is in some ways contradictory. In a study that looked at the correlation between the availability of mobile technology and violence, Shapiro and Weidmann (2012)[43] found that in the case of Iraq, the location of cell phone towers is inversely associated with violence: i.e. that areas of greater access to telecommunications experienced less violence. Using district level data and a difference-in-difference design (a research method for estimating causal effects), the authors find that the expansion of the cell phone network in Iraq is associated with decreases in successful violent attacks by insurgent forces. Shapiro and Weidmann (2013) state that this is due to the extensive use of cell phone surveillance by U.S. and Iraqi anti-insurgent forces as well as successful whistle-blower programs. Similarly, in the African context, Livingston (2011)[44] argues that while cell phones might empower violent groups and produce more violence, there is a potential for a reduction in violence if improved monitoring is done by international peacekeeping or governmental forces. Such efforts have been rare so far, however.

43 Shapiro, J. N. and Weidmann, N. B. (2013). *Is the phone mightier than the sword? Cell phones and insurgent violence in Iraq.* Department of Politics and Woodrow Wilson School, Princeton University. Available from: https://webspace.princeton.edu/users/esocweb/ESOC%20website%20publications/SW_CellphonesIraq.pdf. (Accessed 1 Sept. 2014).
44 Livingston, S. (2011). Africa's Evolving Infosystems: A Pathway to Security and Stability. *Africa Center for Strategic Studies.* Research Paper No. 2.

Alternatively, Pierskalla and Hollenbac (2013)[45] provide evidence to show that cell phone technology can increase the ability of violent groups to overcome collective action problems in Africa. In particular, they state that cell phones lead to a boost in the capacity of groups to communicate and monitor in-group behaviour, thus increasing cooperation. They offer some insights suggesting that the exploration of potential interactions with country or group-level variables can further illuminate the effects of communication technology on violence.

Pierskalla and Hollenbac conclude that enlarging the communication network of violent groups as well as increasing the rate of communication by group members should raise in-group trust between individual participants. The possibility for fast and easy communication boosts the propensity for and rate of information sharing within groups, creating a shared awareness among group members. This system can also be applied to ethnic groups, religious groups or specific sectors of the population. As Shirky (2008, 51) writes, collective action is critically dependent on group cohesion.[46] The expansion of within-group communication is likely to foster shared beliefs and awareness of groups, thus providing one channel of easing collective action. The higher rate of communication between individual group members also makes the transmission of messages and instructions from group leaders through the decentralised network more likely and efficient. Furthermore, the increase in two-way communication vastly raises opportunities for monitoring each other's behaviour (see Sambuli and Awori, p. 29).

Homophily or the closed network effect: a study from the Central African Republic

An example of this phenomenon is currently playing out in the Central African Republic. The Central African Republic has a mobile coverage of 30% and an Internet penetration of 0.1%. Internews is an international non-profit media organisation whose mission is to empower local media worldwide to give people the news and information they need, the ability to connect, and the means to make their voices heard. In the Central African Republic, where Internews has been working since 2010, the organisation works mainly with radio stations – as radio is without any doubt the most widespread medium of communication in the country, and in certain cases, the most trusted.

Even now, when more than 50% of the radio stations have been looted or destroyed, radio remains the only means to broadly reach the local population. But while technology use in the Central African Republic is not yet widespread, certain technology is available and at low cost. A fake Blackberry on the black market costs 15,000 CF (almost 32 USD). Other phones, either with or without Internet capability, cost around 12,000 CF (or 24 USD). Those phones have two things in common: a camera to take pictures and video, and Bluetooth.

In 2014 a new phenomenon emerged in the country: young people were taking video of massacres and killings with their phones to share with friends and peers. Especially in the capital of Bangui, youth began gathering in groups to share videos and pictures of the violence happening in their areas by using Bluetooth, or sometimes by exchanging memory cards. This information flow is a completely closed and untapped one, where access is gained through a shared view of the conflict or geographical and ethnic commonalities. In other words, access to the circle of information comes from having already been a part of it.

This system is typical of the phenomenon of 'homophily', as discussed by Ethan Zuckerman in his closing remarks at the 2014 PeaceTech conference in Boston [47] and increasingly a hallmark of modern conflicts. The homophily principle states[48] that people's personal networks are homogeneous with regard to many socio-demographic, behavioural, and intrapersonal characteristics. Therefore homophily limits people's social worlds in a way that has powerful implications for the information they receive, the attitudes they form, and the interactions they experience. Within this same context, ties between non-similar individuals also dissolve at a higher rate, which sets the stage for the formation of niches (localised positions) within social space. Use of social media and other closed systems for sharing information mean that digital social networks frequently become ways to reinforce views, limit exposure to alternative narratives and thus reduce dialogue and mutual understanding between groups in conflict.

The consequence is that conversations enabled through homophilic systems are more and more polarised towards one unique vision, the vision of the people forming the network. Within the network, the likelihood of a divergent opinion or conversation that

45 Pierskalla, J. H. and Hollenbach, F. M. (2013). Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa. *American Political Science Review*. 107. pp. 207-224. Available from: http://polisci.duke.edu/uploads/media_items/technology-collectiveactioncellphoneviolence.original.pdf. [Accessed 1 Sept. 2014].
46 Shirky, C. (2008) *Here Comes Everybody: The Power of Organizing Without Organizations*. Penguin Press.
47 Video at https://www.youtube.com/watch?v=Mj_SKNQX654. [Accessed 1 Sept. 2014].
48 McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*. 27 (1). pp. 415-444.

presents opposing or different opinions is minimised. The information received in those networks is likely to be aimed at enforcing and supporting a singular point of view, and less likely to be surprising or challenging. Common ground between sides in a conflict is therefore reduced.

This is exactly the situation we observe now in the Central African Republic, where Bluetooth is being used to create a closed information system that can function without Internet and still diffuse information that appeals to people sharing the same 'values' – which may be positive or negative. The use of this system is potentially having huge effects on the behaviours of the local population including acting as an incentive to violence, and is also a possible cause of displacement (see Gilman, p.9; see Grayman and Anderson, pp. 22-26; see also Sambuli and Awori, pp. 27-31). Owing to the lack of vetted and reliable information in the country, the local population makes decisions about its actions based on rumours, fears and word of mouth. The use of mobile phones to spread information that is not only unverified, but can also be manipulated *ad hoc* (for example, showing an old video of a destroyed village and stating that it was just destroyed the day before, therefore increasing the fear and feeling of a continuous attack being perpetrated against one group or another) can further increase the use of non-vetted and non-verified information to make important decisions, like fleeing from a certain area or looking for weapons to prepare for a potential attack.

Digital networks and security management: understanding information flows or controlling them?

The Central African Republic is not the only example of such systems. In 2014 in Kenya, during the armed attack carried out by Al Shabaab fighters at the popular Westgate mall, several messages were circulated via the WhatsApp smart phone application. One of the messages said,

> An intel guy, who is communicating with a military consultant, who is inside Westgate as we speak says that the terrorists are in Barclays Premier with some hostages and shielded by the bullet proof glass. Other hostages are tied to the pillars in the basement with explosives. Suicide bombers have been dispatched to other four unknown locations. Also confirmed that Samantha Lethwaite is the leader.

Another message was also sent over mobile phones,

> Guys, if you know anyone near that area please tell them to move as far as possible! Apparently all of the third and fourth floor are laced with explosives and those guys may blow anytime. Hear there are over a 100 people dead in Nakumatt maybe all or some of the hostages. They are in Nakumatt basement. All hostages surrounded by bombs. So if anyone tries to do anything they will blow it. So they are planning on how to go about it. Message from Special Squad.

Of course none of this information ended up being true or was ever confirmed by the local authorities. However, those messages helped in spreading panic and rumours and fostered an environment of fear and suspicion within the local population. Anecdotal evidence shows that people indeed left their houses and some even the country for fear of possible other attacks or for fear that the Westgate mall might explode. More research is necessary to determine whether any of those actions were indeed caused by the spread of this information over mobile phones (see Grayman and Anderson, pp. 22-26; see also Sambuli and Awori, pp. 27-31).

The messages spread over the WhatsApp application in Kenya have several characteristics in common:

1. The explicit request not to spread the information via social media. This made it impossible to correct, deny or confirm any of the rumours.[49]
2. All messages claim to come from an inside source from the official security apparatus.
3. They were all spread using a closed and existing network, WhatsApp, which is based on personal phone numbers. This means that the messages where spread quickly between people that trusted and knew each other well.

These closed systems, like the one used in CAR, spread quickly and work efficiently because they offer many advantages:

1. They are closed systems and rely on peer to peer trust – I trust you and therefore I trust what you are telling me – which allows for the primary source to become irrelevant to the reliability of the information, because the trust is transferred to others.

49 The capacity of Twitter to generate corrections to rumours was analysed in detail by the London School of Economics following the London riots in 2011. This research found that the power of Twitter users to correct false information was equal to their power to spread it: most rumours were identified as false and corrected within 2-3 hours. See Richards, J. and Lewis, P. (2011). How Twitter was used to spread – and knock down – rumours during the riots. *The Guardian*. 7 Dec. Available from: http://www.theguardian.com/uk/2011/dec/07/how-twitter-spread-rumours-riots. [Accessed 1 Sept. 2014].

2. They allow for the information to spread fast because it is free and relies on homophily; both the Bluetooth and the WhatsApp systems are relatively cheap if not totally free.
3. They prevent any sort of cross-verification from happening. Only people that are inclined to trust the information will receive it and they only share it with others that have their same values, so the likelihood of someone within the system to doubt the information declines considerably.

What these two examples highlight is that technology is not only democratising information but is also sequestering it, confining it into small areas that external actors cannot reach easily, and thereby enabling the creation of more closed systems, rather than open ones. Those systems are based on the existence of confirmation bias, a cognitive stance that favours information that confirms previously existing beliefs.

One of the main differences between the system developed in CAR and the one used in Kenya stems from distinction in the technology used. Systems like WhatsApp, as well as BBM, Twitter and Facebook private conversations, can be monitored by the authorities because they rely on a controlled and accessible infrastructure – the mobile and internet network. Collaborative efforts between authorities and mobile providers have already happened in several cases, such as the London Riots of 2011.[50] On the other hand, systems like Bluetooth are much more difficult to tap into and to monitor because the only way to see what is being exchanged is to have access physically to the phone or to be close enough to the exchange point to tap into it.

The evidence available to date, however, suggests that the approach of controlling or even blocking instant messaging systems has not generated particularly positive effects. Anecdotal evidence on the ground highlights that when a system is not available anymore, people find an alternative to exchange information anyway. No study so far has been able to prove that there are possible beneficial effects deriving from blocking the use of certain technologies. In addition to this, concerns need to be raised in terms of the implications that those types of measures, including surveillance, have when it comes to the right to privacy and to free speech.

There is also a value in being able to understand and see those conversations, and in engaging with the people who take part in them. From a programming and peacebuilding perspective, one of the main possibilities is the opportunity to break the homophily system by inserting voices in the conversation that can bring different and also opposing opinions. From a security perspective there is also a value, albeit an indirect one. As described above, the veracity of the information shared through such networks is often beside the point: therefore, accessing networks of this type is not likely to provide reliable warning of attacks or planned violence *per se*. However, developing ways to track information moving in closed communication systems could provide important insights into the perceptions of conflict and the framework through which parties to conflict interpret events and view those involved (see Grayman and Anderson, pp. 22-26; see also Sambuli and Awori, pp. 27-31).

One very interesting example of a completely different strategy that leveraged homophily and learnt from violent actors for the creation of a peace-keeping and early warning system is a small project implemented in Kenya during the 2013 elections. Sisi Ni Amani,[51] a local organisation, used mobile phones and SMS as a way to intervene in the decision-making processes that led to violence by studying the triggering factors of violence in different contexts. Sisi Ni Amani was able to use existing networks on the ground to develop a strategy that was based on groups' ethnic and demographic affinities. The messages developed and sent by SMS to people identified as vulnerable to violent behaviour were developed and designed by their peers, and therefore built on their common values and confirmation bias.

More applied research in this field is needed. Most of all, however, there is a need to move beyond the above-mentioned dichotomy: technology, along with other tools, can and will be used in positive and negative ways by affected populations. Preventing or blocking the use of certain technologies will not really address the issue. A much deeper understanding of the dynamics of information in conflict and how internal communication flows can be used to increase exposure to 'the other' and opposing views, rather than increase polarisation, is critically needed.

50 Jamieson, D. (2011). London Riots Co-ordinated with BlackBerry Messenger. *TechWeek Europe*. 8 Aug. Available from: http://www.techweekeurope.co.uk/news/london-looting-co-ordinated-with-blackberry-messenger-36303. [Accessed 1 Sept. 2014]. Halliday, J. (2011). London riots: BlackBerry to help police probe Messenger looting 'role'. *The Guardian*. 8 Aug. Available from: http://www.theguardian.com/uk/2011/aug/08/london-riots-blackberry-messenger-looting. [Accessed 1 Sept. 2014].
51 http://www.sisiniamani.org. [Accessed 1 Sept. 2014].

Network analysis, which analyses the relationships and interdependency between interacting units (such as individuals) and is widely used in epidemiology, social anthropology and organisational behaviour, has been used to examine and interpret the dynamics of wars for many years.[52] It has also more recently been applied to understanding how the internet functions, and the same approach – looking at how social networks connect and unite social groups – can be applied to offline systems such as mobile. Moody (2005)[53] suggests that a comprehensive social network analysis can help in identifying the magnitude of social multiplier effects, for example.

We also need to start learning from the use of technology by violent actors. Studying and understanding how already existing systems work can help us understand what they rely on, and leverage this information to create positive counter-systems, much like what Sisi Ni Amani did in Kenya. There is a requirement to look carefully at what is happening on the ground from a more sociological point of view, rather than a security one: in humanitarian emergencies local staff are also affected population, and can offer humanitarian organisations a window into the dynamics and tools used by the local population to communicate.

From a security perspective, it is easy to dismiss the kind of information that moves through closed networks: much of it is clearly (deliberately or accidentally) untrue, or (deliberately or accidentally) misrepresentative of ground realities. Yet the available evidence to date suggests that dismissing this information would be wrong. Such information can be extremely useful in predicting humanitarian problems, such as displacement (in response to rumours or threats), identifying misperceptions (deliberate or accidental) regarding the actions of international agencies, and in understanding the drivers of conflict. Accessing and triangulating this information, however, remains a key challenge.

52 See the work of Emile M. Hafner-Burton, Alexander Montgomery and others.
53 Moody, J. (2005). Fighting a Hydra: A Note on the Network Embeddedness of the War on Terror. *Structure and Dynamics*. 1 (2). Available from: http://escholarship.org/uc/item/7x3881bs. [Accessed 1 Sept. 2014].

Conjuring Zones of Insecurity

Post-Conflict Election Campaigning by Text Message in Aceh, Indonesia

Jesse Hession Grayman / Bobby Anderson

Introduction

Information Communications Technology (ICT)'s systematic penetration into the developing world has fundamentally changed the way people at the margins communicate with one another. The spread of cheap mobile devices has spawned a wave of development initiatives falling under the umbrella of Information Communications Technology for Development (ICT4D). These initiatives are said to be 'creating new venues for people's participation and giving new voice to those who have historically been marginalized.'[54] Mobile devices now allow hitherto excluded communities to access banking services in Kenya,[55] report corruption in India,[56] improve health services in Peru[57] and Rwanda,[58] raise educational levels in South Africa[59] and Tanzania,[60] and monitor elections and report electoral violence in Kenya,[61] Nigeria[62] and Indonesia.[63] Estonia even allows citizens to vote through mobile devices.[64] Mobile devices also powered the 'Arab Spring' on the ground in Egypt, Libya and Tunisia.

These examples may or may not, in the clichéd benchmark of development, prove 'sustainable' over time, but they do represent concrete examples of ICT4D's transformational possibilities. We would insert an important caveat that ICT tools bear the moralities of their wielders. In this collection, Sambuli and Awori (pp. 27-31) discuss how mobile and digital technologies may have played a catalysing role in the Kenyan 2007/08 post-election violence, and how the Umati project was created to track and counter such use in 2013.

In Afghanistan the night-letter phenomenon of the early occupation has moved from physical letters to text messages (SMS, Short Message Service) across cell phone networks. ISIS in Iraq and Syria embraced mobile technologies more than the nascent democrats in any preceding Arab Spring. Where we work in Indonesia, mobile technologies have been widely embraced in politics and elections, especially among gangster, police, military, and militia groups, to intimidate and manipulate outcomes. A differentiated spectrum of ICT is deployed across the archipelago: in Jakarta, youth gangs taunt one another and arrange fights and ambushes via Twitter.[65] In Maluku, e-rumour-mongering in the ethno-religious fracture zones of urban Ambon leads to arson, assaults and murders.[66] In Aceh, threats via SMS were readily adopted to extort from contractors in the post-tsunami reconstruction boom, but the use of SMS in that province has had more widespread impact through the cultivation of pervasive and ephemeral environments of insecurity that arrive with every election cycle. Most frustrating is that while these SMSs add to a feeling of insecurity that can penetrate uninvolved households, there is little concrete action that can be taken against them (see Ayala, pp. 19-21; see also Sambuli and Awori, pp. 27-31).

The quarter-century conflict between the Government of Indonesia and the Free Aceh Movement (known by their Indonesian acronym GAM, from Gerakan Aceh Merdeka) was distinguished by sporadic separatist violence and brutal counter-insurgency operations.

54 UNDP. (2012). Mobile Technologies and Empowerment: Enhancing human development through participation and innovation. Available from: https://www.undpegov.org/mgov-primer.html. [Accessed 1 Sept. 2014].
55 Karugu, W. N. and Mwendwa, T. (2007). Vodafone and Safaricom Kenya: Extending the Range and Reliability of Financial Services to the Poor in Rural Kenya. UNDP. Available from: http://growinginclusivemarkets.org/media/cases/Kenya_MPESA_2008.pdf. [Accessed 1 Sept. 2014].
56 http://ipaidabribe.com. [Accessed 1 Sept. 2014].
57 http://healthmarketinnovations.org/program/nacer. [Accessed 1 Sept. 2014].
58 UN. (Undated). Tracnet, Rwanda: Fighting Pandemics through Information Technology. Available from: http://www.un.org/esa/sustdev/publications/africa_casestudies/tracnet.pdf. [Accessed 1 Sept. 2014].
59 http://www.comminit.com/africa/content/dr-math. [Accessed 1 Sept. 2014].
60 Trucano, M. (2009). Checking in with BridgeIT in Tanzania: Using mobile phones to support teachers. World Bank. 25 Sept. Available from: http://blogs.worldbank.org/edutech/checking-in-with-bridgeit-in-tanzania. [Accessed 1 Sept. 2014].
61 http://www.ushahidi.com. [Accessed 1 Sept. 2014].
62 http://www.reclaimnaija.net/
63 http://www.kawalpemilu.org. [Accessed 1 Sept. 2014].
64 ICT Statistics Newslog. (2011). Estonians Vote in Parliamentary Election by Mobile Phone. 7 March. Available from: http://www.itu.int/ITU-D/ict/newslog/Estonians+Vote+In+Parliamentary+Election+By+Mobile+Phone.aspx. [Accessed 1 Sept. 2014].
65 Anderson, B. and Snyder, J. (2013). Coming of Age in the Urban Kampung: Gang Demographics and Territories in Select Jakarta Neighborhoods – Preliminary Findings. A paper presented at the European Association for South East Asian Studies Conference, Lisbon, Portugal, 2-5 July 2013. Available from: https://www.academia.edu/3849904/Coming_of_Age_in_the_Urban_Kampung_Gang_Demographics_and_Territories_in_Select_Jakarta_Neighborhoods-_Preliminary_Findings. [Accessed 1 Sept. 2014].
66 Spyer, P. (2002). Fire without smoke and other phantoms of Ambon's violence: Media effects, agency, and the work of imagination. *Indonesia*. 74. pp. 21-36. Available from: http://cip.cornell.edu/DPubS?service=UI&version=1.0&verb=Display&handle=seap.indo/1106939683. [Accessed 1 Sept. 2014].

The conflict ended in August 2005, when the Helsinki peace agreement was signed, catalysed in part by the promise of enough lucrative spoils for all former adversaries in Aceh's post-tsunami reconstruction economy. By mid-2008, former GAM separatists were preparing to contest seats in Aceh's provincial and district assemblies for the first time with their newly formed local political party, Partai Aceh (PA). The allowance for Aceh to field local parties in the April 2009 legislative elections was the first of its kind in Indonesia, and a benchmark achievement for GAM in the peace agreement.

Polarising the electorate

In mid-2008, while one of us (JHG) conducted ethnographic field research and the other (BA) implemented reintegration and stabilisation projects for post-conflict recovery in Aceh, we noticed that interviewees and other local partners in our work would routinely take out their cell phones to show us the frequently anonymous, political text messages they had received to illustrate ongoing tensions among conflict-era adversaries. One of the most memorable text messages was written in traditional Acehnese verse and sent anonymously to the cell phones of a select group of ex-GAM rebels who were not on the best of terms with their former comrades now campaigning for Partai Aceh:

> A YOUNG CHILD GATHERS RATTAN IN THE MOUNTAINS OF MEUREUDU / FIND THE BEST TO MAKE A BASKET / NOW IT IS ALMOST ELECTION SEASON / IT IS TIME TO CHOOSE A THRONE FOR THE KING / HEAD OVER THERE TO GAM'S PARTY / HAVE NO DOUBTS MY BROTHER / WHOEVER DOES NOT CHOOSE THE DESCENDANTS OF ACEHNESE KINGS / JUST MOVE TO JAVA / NO NEED TO STAY ANYMORE IN ACEH / JUST GET THE FUCK OUT OF HERE[67]

Recipients of this poetic intimidation were all GAM ex-combatants who surrendered before the peace agreement and underwent formal reeducation sponsored by the Indonesian military (Tentara Nasional Indonesia or TNI); the larger GAM conglomerate thus considers them traitors. During the final years of the conflict, these reformed ex-rebels operated as any of the other anti-separatist militia groups in Aceh with TNI backing, and in the early post-conflict era were seen as potential spoilers of the peace process. Not to be outdone, they expressed their disappointment with GAM's leaders by widely distributing an SMS of their own:

> IN THE YEAR 2000 WE RAN AWAY, FEARFUL OF POLICE AND SOLDIER'S WEAPONS. IN THE YEAR 2004 THE TSUNAMI CAME, ALLAH'S JUDGMENT THAT BROUGHT ENORMOUS WATER. IN THE YEAR 2006 THERE WAS NO MORE FIGHTING. IN THE YEAR 2007 WE INAUGURATED NEW KINGS. IN THE YEAR 2008 THEY FOUGHT AMONGST THEMSELVES. THE LEADERS OF THE LAND FORGOT TO COMPENSATE THEIR PEOPLE'S SERVICE. NOBODY CARES ABOUT THE VICTIMS OF SHOOTINGS, NOR DOES ANYBODY CARE ABOUT THE WIDOWS. THE ARISTOCRATS AND DISTRICT LEADERS ARE BUSY WITH THEIR TOYOTA LUXURY VANS. IN THE YEAR 2009 WE CHOOSE THE PEOPLE'S REPRESENTATIVES, AND AGAIN THEY BRING US PROMISES ON A HEAVENLY WIND. THOSE PROMISED A CAR WILL GET A BICYCLE. THOSE PROMISED A COFFEE WILL GET POISONED. CONGRATULATIONS TO THE LEADERS OF THIS LAND!

This exchange evokes the simmering tensions between ex-combatants, with implied threats from the first SMS and disappointment expressed by the second (a disappointment that has taken root across Aceh since those elections). Meanwhile, TNI fixated on the possibility of a resurgent separatist threat if PA won the elections. One officer at a base in East Aceh sent the following SMS to village heads in neighbouring sub-districts:

> BE CAREFUL, GAM HAS BEGUN LISTING YOUR CONSTITUENTS AS MEMBERS OF THEIR POLITICAL PARTY BY FILLING IN BLANK GAM PARTY FORMS. GAM'S METHODS ARE NOT SO DIFFERENT FROM THOSE USED BY THE PKI [THE INDONESIAN COMMUNIST PARTY] IN THE PAST. DO NOT BE SEDUCED BY GAM'S DECEPTION; IT COULD BE A TRAP, BUT IF PEOPLE WANT TO THEN FEEL FREE TO FILL IN THE FORMS COMPLETELY. SHARE THIS SMS WIDELY WITH YOUR FAMILY, NEIGHBOURS, FRIENDS, ETC., SO THAT PEOPLE IN THE COMMUNITY ARE NOT DECEIVED, AND BECOME VICTIMS LIKE THOSE CAUGHT UP IN THE PKI'S SEPTEMBER 30TH MOVEMENT REBELLION IN 1965.

67 The translations here and below are by JHG. The final verse in this message uses an unspeakably rude metaphor in Acehnese that our colleagues in Aceh insisted has no equivalent English meaning. The translation conveys the threat to leave Aceh with an equally offensive English expression.

The village head who showed it to us could not tell if the officer who sent it composed it himself or if he was simply forwarding it from another source. The message is a thinly veiled threat suggesting that the fate of PA members may resemble the fate of communists in 1965 who were massacred in the hundreds of thousands across Indonesia, including Aceh. Messages linking the fate of those who support GAM to those who were in the PKI were common in Aceh at the time, and were not limited to elections; in some reintegration programs that targeted ex-GAM in Aceh's highlands, local TNI circulated SMS messages stating that the GAM who signed up for such programs were also signing their names to 'death lists.' Anti-separatist militias in the highlands also sent out messages with similar themes.

Messages like these do not typically appear in the mass media or in analyst reports about post-conflict politics in Aceh, and yet our data and experience show that this global technology was routinely deployed to spread rumours and threats, campaign promises and political slander, poetry and invective, all across the province, often in rich Acehnese and Indonesian vernaculars. SMS technology may be the most cost-effective election campaign tool because it penetrates remote communities without requiring travel, and reaches voters and adversaries much more reliably and cheaply than telephone, radio, or television broadcasts (see Ayala, pp. 17-18). The medium also allows for anonymity; senders can terrorise individuals and communities from a distance. And they frustratingly provide little to act against. Most of these ephemeral documents transmitted across cell phone networks easily escape the archives that bear only partial historic witness to what was a momentous and occasionally tumultuous transition to peace.

Intimidation across phone networks during the 2009 elections

During the March 2009 Aceh legislative elections, one of us (JHG) worked as an election observer with an international NGO. As pairs of observers moved from one town to the next, our contact information travelled through election stakeholder networks, and we soon found it difficult to accommodate, much less sort out and make sense of, the barrage of data that people sent us via SMS, frequently from unknown sources. A few examples recall the atmosphere of fear that voters, candidates, officials, and other election stakeholders conveyed to us:

> YOU CAN SEND OUR BROTHER TO JAIL, BUT I WILL SEND YOU, COMMISSIONER, TO THE GATES OF HELL. GO AHEAD, ENJOY YOUR LIFE WITH YOUR WIFE AND CHILDREN, ONLY A FEW MORE MOMENTS REMAIN. [SENT TO THE HEAD OF THE BENER MERIAH DISTRICT ELECTION COMMISSION, FORWARDED TO US.]

> SIR, DO NOT RETURN SO FREQUENTLY TO YOUR HOME. WHEN YOU RETURN HOME, YOU'LL BE SHOT DEAD IMMEDIATELY. THIS IS VALID INFORMATION. WE HAVE THE WEAPONS NEAR SIMPANG MAMPLAM. [MESSAGE SENT TO A LOCAL PARTY CANDIDATE, FORWARDED TO US].

> ON TUESDAY NIGHT AT AROUND 1:30AM, SIX OFFICERS FROM THE TNI BASE ARRIVED ON THREE MOTORBIKES, CARRYING THREE FIREARMS. THEN THEY TOLD US 'DO NOT VOTE FOR ACEH. IF YOU VOTE FOR ACEH IT MEANS YOU'RE INVITING WAR WITH US.'

> THE TAMIANG POLICE CHIEF AND HIS MEN HAVE SURROUNDED THE HOME OF THE DISTRICT HEAD OF PARTAI ACEH, AND WE DON'T KNOW WHY. THE INTIMIDATION HERE IS SEVERE. PLEASE INVESTIGATE AND RESPOND.

> GOOD EVENING, WE ARE VERY FEARFUL OF THE TNI AND THE POLICE WHO HAVE BEEN ROAMING ABOUT. AT NIGHT, THEY ARE EVERYWHERE LIKE OWLS, BUT IN THE DAYTIME THEY DO NOT APPEAR. WE ASK THAT YOU WILL PUBLICISE THIS INFORMATION IN THE INTERNATIONAL NEWS, THAT THE PEOPLE OF ACEH ARE AFRAID OF THE TNI AND THE POLICE. PLEASE DO NOT SHARE MY PHONE NUMBER WITH ANYONE. THANK YOU.

These messages are anecdotal examples of overlapping individual and group communication strategies that were finding their way into the majority of mobile phones in Aceh. The volume of election-related SMS is not quantifiable, but our experience indicates that it was overwhelming. Intimidations by text message generally had their intended effect as they circulated in a setting of pre-election violence including arson, bombs, and targeted murders. Text messages reporting these violent acts (and threatening more to come) produced a sense of immediacy and proximity, amplifying and personalising among ordinary voters the effects of what were mostly isolated and parochial clashes between distant adversaries.[68] Rival candidates took to sleeping at different houses each night. They, and ordinary citizens, also ensured they were off the road before nightfall.

Intimidation by SMS: a winning campaign strategy?

Just five days before the election, in the early evening on 4 April, JHG was meeting with a local NGO at a popular restaurant in the city of Langsa, and it was there that the atmospherics of terror turned into concrete reality when his interviewee received an urgent SMS:

> TEUNGKU LEUBE, THE FORMER REGIONAL [GAM] COMMANDER OF ARAMIAH LANGSA [JUST WEST OF LANGSA IN EAST ACEH] AND CURRENT HEAD OF THE PA SUB-DISTRICT OFFICE THERE, AGE 41, HAS BEEN SHOT DEAD BY UNKNOWN ASSAILANTS AT AROUND 7:20PM. HIS BODY HAS BEEN BROUGHT TO THE LANGSA PUBLIC HOSPITAL.

Moments later, similar messages arrived on JHG's phone as well, with PA officials requesting the election observers bear witness. Five minutes later JHG arrived at the hospital, where a large crowd already stood outside the emergency room. This was the sixth murder of a PA activist since February 2009.

Despite these terrors visited upon PA activists, on election day PA easily won nearly half the votes province-wide, allowing them to dominate the provincial assembly. For their part, during the campaign PA had directed plenty of their own threats and intimidation, mostly toward the other five local political parties, ensuring only PA took part in organised local politics. In Aceh's subsequent 2013 gubernatorial and 2014 legislative elections, PA managed to unseat the enormously popular incumbent governor by creating an environment of pervasive threat not unlike what they had experienced in 2009. The number of violent incidents in the months immediately preceding the 2013 election correlated with the shift in public support to PA's ticket.[69] These former separatists threatened to return the province to war if they lost, and on that hopeful platform, they won. In the run-up to Aceh's 9 April 2014 legislative elections, the Commission for Missing Persons and Victims of Violence (Kontras) counted 48 cases of election-related violence from January to March, including murder. In both elections, SMS threats served as force multipliers for numerous targeted killings and bombings: intimations in their arrival on an uncounted and unsolicited number of mobile phones that no one was safe from such outcomes.

Conclusions

Mobile technologies have been as thoroughly embraced by ex-insurgent thugs, religious extremists, and other 'uncivil' societal forces as they have by Kenyan housewives, Tanzanian cattle traders and Indonesian electoral quick-count monitors. A decade ago in Indonesian Papua, when select military units wished to clear Jayapura's evening streets, they would circulate rumours of vampires killing children in major towns. These actors learned that SMS cheaply and efficiently cultivate and expand zones of insecurity; the last decade has shown the nefarious embrace of this seemingly innocuous tool.

One counter-response from peace activists includes the development of 'early warning, early response' methodologies with ICT elements, often using the same mobile technologies first employed by spoilers in pursuit of more violent ends. International Crisis Group has reported on a loosely organised group, the 'peace provocateurs,' that responded to re-emergent sectarian violence in Ambon, where partisans have long relied upon media technologies to amplify the atmospherics of insecurity. The peace provocateurs now respond with the same tools, bringing their mobile phones to trouble spots to check rumours, then send back information and photos to a point

68 On immediacy and proximity, see: Grayman, J. H. (2014). Rapid response: Email, immediacy, and medical humanitarianism in Aceh, Indonesia. *Social Science & Medicine*. In press. http://dx.doi.org/10.1016/j.socscimed.2014.04.024. [Accessed 1 Sept. 2014].
69 Anderson, B. (2013). Gangster, ideologue, martyr: The posthumous reinvention of Teungku Badruddin and the nature of the Free Aceh Movement. *Conflict, Security & Development*. 13 (1), p. 53.

person who broadcasts updates to journalists and the wider public via SMS.[70] This strategy, however, is a reactive solution, a stop-gap measure always one step behind potential spoilers.

Our experiences with these proliferating mobile technologies throughout the past decade in Aceh and elsewhere in Indonesia have led us to consider their security implications for aid and development projects. Text messages are now the first medium of choice in threatening and extorting not just stakeholders in local elections, but also aid projects and staff involved in the broad milieu of conflict recovery efforts.

The phenomenon of SMS-driven threats and rumours does not generally interfere with the sharing of genuine security information amongst security planners. As disseminators of rumour or threat, they are as common in deteriorating environments as confetti (a better comparison may be the chaff that disrupts flight radar), and reacting to them all is neither necessary nor possible: JHG's experience shows that they can contain relevant information, which is to be noted, but not responded to unless particularly rich in plausible detail. The overall security environment influenced how affected people made personal security decisions: political and ex-insurgent notables changed their travel and accommodation patterns in response to this, not driving on the same routes frequently, even relocating temporarily to safer and less-known locations, sometimes away from their families, or ensuring a security presence in their homes. Threatening SMSs generally served as symptoms of the environment, rather than anything actionable in themselves.

The vast majority of threatening SMSs that BA received while he managed dozens of stabilisation and reintegration projects in Aceh did not escalate beyond his phone and ended with the delete button. Only on three occasions in 3.5 years did SMS threats eventually lead to actual violence. Ease of delivery and the ability to change numbers suggests that these threats should not be taken as seriously as traditional ones (see Sambuli and Awori, pp. 27-31). But security managers should establish protocols for this medium: logging messages and numbers used, with particular attention to any personal information directed at individuals. Non-specific threats that contain no intimate information (such as addresses or license plate numbers or names of children or spouses, for example) should be treated as anecdotal measurements of the increasing volatility of a security environment, and it is that environment that security managers will react to, rather than the SMS. SMSs containing specific information, however, need to be reacted to, as do SMSs arriving on, for example, an individual's private number which they do not disseminate widely (implying the threat's origin is more intimately close to the individual).

Over-reaction to non-specific threats weakens project credibility in post-conflict contexts precisely because such threats are to be expected in the course of such work. Numerous NGOs in Aceh actually used the threat environment to declare *force majeure* on projects that were failing long before they received any SMS threats. The urge to respond to such SMSs also needs to be resisted, because generally a response indicates that the threat has made an impact, and exposes the respondent to more threats, and more manipulation.

In conclusion, SMS threats do not generally constitute threats in themselves. Their anonymity and ease of use allow them to take on the appearance of threats, but they generally only act as the connective tissue for a larger body of more obvious threats that persons and organisations can act against. With certain exceptions, they serve as indicators for insecurity, rather than the insecurity itself.

70 International Crisis Group. (2012). Indonesia: Cautious Calm in Ambon. *Asia Briefing*. 133. 13 Feb. p. 2. Available from: http://www.crisisgroup.org/en/regions/asia/south-east-asia/indonesia/b133-indonesia-cautious-calm-in-ambon.aspx. (Accessed 1 Sept. 2014). See also ICG's Asia Briefing 128, which first reports on the formation of the peace provocateurs.

Monitoring Online Dangerous Speech in Kenya

Insights from the Umati Project

Nanjira Sambuli / Kagonya Awori

Introduction

The Umati[71] project emerged out of concern that mobile and digital technologies may have played a catalysing role in the Kenyan 2007/08 post-election violence, and that the online dissemination of potentially harmful speech was inadequately monitored. In the build up to the 2007 Kenyan elections, avenues of propagating dangerous speech were generally limited to broadcast media transmissions, print media, SMS and email. Anecdotal evidence suggested that online spaces such as forums and blogs were also used to plan and incite violence on the ground. However, at that time, no system existed to track such data. Incendiary remarks by politicians and notable public figures such as musicians (through lyrics) have been noted to incite violence in Kenya's historical past, specifically around election periods, with a culmination noted during the 2007 election period and its aftermath. Efforts to monitor hate speech have been in place through undertakings by Kenyan civil society as well as police authorities. However, the migration of inflammatory speech to online media remained neither monitored nor analysed.

Since the submarine fibre optic cables landed on Kenyan shores in 2009, Internet penetration has been on a steep increase.[72] Greater access to affordable Internet, especially through the use of smart and feature phones,[73] has seen increased use of social media in the country. Such platforms offer new spaces for people to express their opinions, especially during times of heightened anxiety such as election periods. With over 2 million active[74] Kenyan Facebook users as of April 2013[75] (an estimated 19.2% of the country's online population), and over 2.48 million geo-located tweets generated in Kenya in the 4th quarter of 2011,[76] it can be deduced that social media is heavily used by Kenyans, and will continue to grow in popularity.

New media have diversified the audiences that engage in online communication. As these online spaces are a new medium for disseminating inflammatory speech, their influence on the actions of the audience warrants assessment. A possible result is the creation of a vicious cycle as audiences convene around hateful content, converse in self-selected groups and form new ideas or support their original biases with the hateful beliefs of others (see Ayala, pp. 17-21). However, there is a prospect of virtuous cycle creation, as new media spaces can also act as alternative information sources that neutralise the negative impacts of online and offline inflammatory speech. An example of this is noted in the findings section below.

Initially, the Umati project sought to better understand the use of dangerous speech in the Kenyan online space by monitoring particular blogs, forums, online newspapers, Facebook and Twitter. Online content monitored includes tweets, public status updates and comments, posts and blog entries. Umati was launched in October 2012, six months before the Kenya general elections (March 4, 2013) and exists in two distinct phases.

Phase I (September 2012 to May 2013) established the following initial goals:

- To monitor and understand the type of online speech most harmful to Kenyan society.
- To forward calls for help to Uchaguzi, a technology-based system that enabled citizens to report and keep an eye on election-related events on the ground.[77]

71 Umati is Kiswahili for 'crowd'.
72 Communications Commission of Kenya. (2013). *Quarterly Sector Statistics Report: First Quarter of The Financial Year 2013/14 (Jul-Sept 2013).* Available from: http://ca.go.ke/images/downloads/STATISTICS/Sector%20Statistics%20Report%20Q1%202013-14.pdf. (Accessed 2 Sept. 2014).
73 *Ibid.*
74 Number of people active on Facebook over a 30-day period.
75 Social Bakers Statistics. (2013). Available from: http://www.socialbakers.com/. (Accessed 12 April 2013).
76 Portland Communications. (2012). New Research Reveals How Africa Tweets. 1 Feb. Available from: http://www.portland-communications.com/2012/02/new-research-reveals-how-africa-tweets. (Accessed 2 Sept. 2014).
77 Uchaguzi was an election-specific deployment by Ushahidi and other stakeholders that saw collaboration between citizens, election observers, humanitarian response agencies, civil society, community-based organisations, law enforcement agencies and digital humanitarians to monitor elections.

- To define a process for online dangerous speech tracking that could be replicated in other countries.
- To further civic education on dangerous speech, and sensitise the Kenyan public in order that they are more responsible in their communication and interactions with people from different backgrounds.

Phase II (July 2013 to January 2016) further aims:

- To refine the Umati methodologies developed in Phase I and where applicable, increase scalability of the project through automation.
- To test the Umati methodology in other countries in order to improve and increase its global/contextual applicability.
- To explore non-punitive, citizen-centred approaches for reducing dangerous speech online.

Umati methodology for identifying dangerous speech

Umati uses Susan Benesch's definition of dangerous speech, that is, speech that has the potential to catalyse collective violence.[78] Benesch's 'Dangerous Speech Framework' offers the following key variables for identifying dangerous speech:[79] the speaker and his/her influence over a given audience – a political, cultural or religious leader or another individual with a significant following tends to have more influence over a crowd; a vulnerable audience subject to incitement by the influential speaker; the content of the speech that may be taken as inflammatory to the audience and be understood as a call to violence; the social and historical context of the speech – for instance, previous clashes or competition between two groups can make them more prone to incitement; and the medium of disseminating the speech, including the language in which it was expressed.

Umati built on the Benesch framework to form a practical identification method. Specifically, the project found that the following three components of the framework were the most relevant for the identification of online dangerous speech in Kenya:[80]

1. It targets a group of people. It is important to note that a hateful comment about an individual is not necessarily dangerous speech unless it targets the individual as part of a group. In our research, it was observed that dangerous speech towards a group can occur across various lines, including religion, tribe/ethnicity, gender, sexuality, political affiliation and race.
2. It may contain one hallmark of dangerous speech. Three hallmarks that are common in dangerous speech comments, as identified by Susan Benesch,[81] include:

   a. Comparing a group of people with animals, insects or vermin;

   b. Suggesting that the audience faces a serious threat or violence from another group, specifically the same group that is a target of the inflammatory speech ('accusation in a mirror'); or

   c. Suggesting that some people from another group are spoiling the purity or integrity of the speakers' group.
3. It contains a call to action. Dangerous speech more often than not encourages a particular audience to commit acts of violence towards a group of people. These can include calls to kill, beat/injure, loot, riot, and forcefully evict.

Umati Phase I relied on a manual process of collecting and categorising online dangerous speech. Human input proved necessary for contextually analysing and categorising speech statements, which in turn facilitated the creation of an inflammatory speech[82] database. Between October 2012 and November 2013, up to eleven monitors scanned a collection of online sites in seven languages: English and Kiswahili (Kenya's official and national languages respectively); Kikuyu, Luhya, Kalenjin and Luo (vernacular languages from the four largest ethnic groups); Sheng (a pidgin language incorporating Kiswahili, local languages and English); and Somali (spoken by the largest immigrant community).[83] In Phase II, the Umati team has begun work on incorporating more automation in the data collection process where applicable. This is being explored through Machine Learning and Natural Language Processing techniques and tools, which if successful, will significantly increase the scalability and transferability of the Umati project going forward.

78 Benesch, S. (2013). Dangerous Speech: A Proposal to Prevent Group Violence. 23 Feb. Available from: http://voicesthatpoison.org/guidelines. [Accessed 2 Sept. 2014].
79 Not all variables must be present for speech to qualify as dangerous speech. Variables are also not ranked and may carry more or less weight depending on the circumstances. Each instance of speech must be evaluated in terms of the information available.
80 For further analysis see Awori, K. (2013). *Umati Final Report: September 2012-May 2013*. p. 27. Available from: http://www.research.ihub.co.ke/uploads/2013/june/1372415606__936.pdf. [Accessed 2 Sept. 2014].
81 Benesch, S. (2008). Vile Crime or Inalienable Right: Defining Incitement to Genocide. *Virginia Journal of International Law*. 48 (3). Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1121926. [Accessed 2 Sept. 2014].
82 Inflammatory speech is used to refer to all three categories along the continuum: offensive speech, moderately dangerous speech and extremely dangerous speech.
83 The sources list currently covers 80+ blogs and forums, 350+ Facebook users, groups and pages, 400+ Twitter users, all major online Kenyan newspapers and YouTube channels for the five main Kenyan media houses.

Categories of inflammatory speech and their likelihood to catalyse violence

As monitors manually scanned online platforms for incidents of dangerous speech, they recorded the speech acts they perceived to be hateful in an online database. In this process, all dangerous speech statements were translated into English and sorted into three categories (in ascending order of severity):[84]

1. **Category one – offensive speech:** mainly insults to a particular group. Often, the speaker has little influence over the audience and the content is barely inflammatory, with no calls to action. Most statements in this category are discriminatory and have very low prospects of catalysing violence.
2. **Category two – moderately dangerous speech:** comments are moderately inflammatory and made by speakers with little to moderate influence over their audience. Audiences may react differently; to some, these comments may be highly inflammatory, while to others, they may be considered barely inflammatory.
3. **Category three – extremely dangerous speech:** statements are made by speakers with a moderate to high influence over their audience. These statements are seen to have the highest potential to inspire violence, as they tend to constitute an action plan that can be understood and acted upon by the targeted audience. These statements are often stated as truths or orders. Umati categorised all statements with a clear or perceived call to beat, to kill, and/or to forcefully evict a particular group, or an individual because of their belonging to a particular group, as extremely dangerous speech statements.

It is important to note that a causal link is almost impossible to draw between dangerous speech and on-the-ground violence, due to the many factors that contribute to bringing about a physical violent act (see Grayman and Anderson, pp. 25-26). However, speech has the capacity to catalyse or inflame violence. Actors are still legally and morally responsible if they commit violence in response to incitement or dangerous speech. When imminent threats of violence were found during the election period, the Umati team extracted the relevant information and forwarded it by email to a listserv of specific stakeholders. These included donor agencies, Umati partners and Uchaguzi key decision makers who were better equipped with mitigating the threats of violence that Umati collected. This process was triggered five times from January to April 2013 and on-the-ground teams mobilised based on the information passed to them.

Findings

- Over 90% of the inflammatory speech statements that Umati collected in 2013 were from Facebook. This has been attributed to the fact that Facebook is the most popular social media site in Kenya. Umati found however, that other factors come into play that accommodate dangerous speech on Facebook as opposed to the second most popular social media site, Twitter.[85] Most interestingly, a behaviour one of the authors named 'KoT cuffing' was observed on Twitter where '[offensive] tweets not acceptable to the status quo are shunned, and the author of the tweets, is publicly ridiculed. The end result is that the "offender" is forced to retract statements due to the crowd's feedback, and can even close his/her Twitter account altogether.'[86] KoT cuffing, a self-policing behaviour by Kenyans On Twitter (KoT), demonstrates that netizens themselves are capable of employing non-judicial means to counter online dangerous speech.
- Not surprisingly, it was possible to identify those that engaged in dangerous speech online, either via their real names, e.g. by use of their Facebook and Twitter accounts, pseudonyms which can be mapped to their email addresses, or through a traceable history of online activity using tracking software. Umati, however, did not attempt to uncover the true identities of online speakers, and instead focused on observing behavioural patterns of repeat dangerous speech offenders over short periods of time.
- Umati data reflected that in Kenya, ethnicity is a primary lens through which political, economic and social issues are viewed and reacted to by the public. Umati data showed that online discriminatory speech is mostly along ethnic lines. However, as different events transpired through 2013, most notably the Nairobi Westgate Mall attack,[87] Umati data shows that Kenyan online discriminatory speech has escalated along ethno-religious lines. What is crucial to note here is not that discrimination is mostly ethnic or religious,

84 Awori, K. (2013). *Umati Final Report: September 2012-May 2013*. p. 27. Available from: http://www.research.ihub.co.ke/uploads/2013/june/1372415606__936.pdf. [Accessed 2 Sept. 2014]. The full categorisation formula, including the data entry form, can be viewed in the final report.
85 Further discussed in Awori, K. (2013). *Umati Final Report: September 2012-May 2013*. pp. 24-25. Available from: http://www.research.ihub.co.ke/uploads/2013/june/1372415606__936.pdf. [Accessed 2 Sept. 2014].
86 *Ibid.*
87 Daily Nation. (2013). Security forces move to end Westgate mall siege as death toll rises to 62. 23 Sept. Available from: http://www.nation.co.ke/news/Westgate-Mall-attack-alshabaab-terrorism/-/1056/2004630/-/kr74w0/-/index.html. [Accessed 1 Sept. 2014].

but that such discrimination often stems from political, economic and social tensions along various divides. Thus, analysis of dangerous speech should be put into context of other speech online, as rarely do such speech incidents happen in isolation. Moreover, efforts to tackle dangerous speech should focus on addressing the deeper-seated issues that drive people to engage in, disseminate and even act on such speech's provocations.

- While the languages used to disseminate dangerous speech are those that are widely understood in the country, Umati collected some instances of coded language that had been used in past election periods. Additional research is required to investigate this linguistic 'code-switching', which is when a speaker alternates between two or more languages in the context of a single conversation, often to convey a thought or say something in secret.
- Umati Phase II has taken a keen focus on counter-speech, based on emerging phenomena on how 'netizens' are dealing with inflammatory speech online. Umati is monitoring how public conversations take place online over time, how some of these conversations may move towards dangerous speech, and the resultant counter-speech efforts if any. This broader approach will help us better understand self-regulation mechanisms employed by online communities (see Ayala, pp. 17-21). Preliminary self-regulation mechanisms observed online include ridiculing a speaker or a narrative that attempts to inflame hate/misinform/disinform, e.g. the aforementioned KoT cuffing; flooding online spaces with positive counter messages that diffuse tensions arising from hateful messages; and the use of humour and satire to 'hijack' inflammatory narratives.

Conclusions

Observations of dangerous speech should be framed within the context of other conversations online, as inflammatory speech statements rarely happen in isolation. Online dangerous speech is a symptom of the much more complex offline socialisations and perceptions that precede online interaction. We are yet to find concrete instances of online dangerous speech catalysing events offline (see Grayman and Anderson, pp. 22-26). Nonetheless, as 'netizens' congregate and converse online, forming networks around issues of interest, the possibility of organising offline reactions to online conversations is likely.

As part of our third objective in Phase II, we will explore efforts to reduce online dangerous speech through online and offline civic engagement. Umati intends to engage with relevant stakeholders on matters pertaining to freedoms of speech and expression towards better understanding how these are understood and exercised by the Kenyan public. While we are primarily looking at online methods, we will build on experience from *NipeUkweli*[88] (Kiswahili for 'Give me truth'), which is an outreach campaign fashioned to explore proactive ways of mitigating dangerous speech both online and offline.

Going forward, we offer that findings from Umati can provide insight into how humanitarian NGOs can galvanise their crisis prevention efforts and help manage security risks, before and during highly polarised events such as general elections (see Grayman and Anderson, pp. 22-26). One possible avenue could be to promote fissures and spaces where citizens in conflict-prone areas can air out any misconceptions or grievances that would otherwise inform hate/inflammatory/dangerous speech, and even violence.[89] Efforts to tackle dangerous speech (and its consequences) should focus on addressing the deep-seated issues that drive people to engage in, disseminate and act on the provocations of such speech.

88 Njeru, J. N. (2013). *NipeUkweli: Outreach to Sensitize Communities on Dangerous Speech: Summary Report*. iHub Research. 20 March. Available from: http://www.ihub.co.ke/ihubresearch/jb_NipeUkweliSummaryReportMarchpdf2013-11-18-16-07-39.pdf. [Accessed 2 Sept. 2014].
89 A creative example of this is the 'Alternatives to Violence Program', in countries like Kenya and Rwanda: http://www.avpkenya.org. [Accessed 2 Sept. 2014].

From Kenya to Myanmar

Though Umati's methodology was designed to monitor online dangerous speech in Kenya, the project's methodology was adopted in early 2014 for a pilot study of online dangerous speech in Ethiopia. Various elements of the coding form were edited to suit the Ethiopian context.[1] Overall, the methodology was applicable and the same categorisation of dangerous speech into three spectra was employed.

Umati is currently piloting the project in Nigeria, ahead of the 2015 elections. We are working with local Nigerian civil society organisations, offering technical support, as the teams adopt the methodology for their context. The Umati team was also recently in Myanmar, sharing insights on setting up the project with civil society organisations such as MIDO[2] who are keen on monitoring and countering dangerous speech online. As the collection and analysis process continues to be improved in Kenya, the aim is that the methodology will remain explicit enough to be understood and redesigned for other country contexts. Findings drawn from Umati's experience in Kenya can guide organisations in managing risks in contexts where online media is a possible vehicle for catalysing dangerous speech and violence.

For further information on the Umati project, see http://www.ihub.co.ke/umati

1 *Gagliardone, I., Patel, A. and Pohjonnen, M. (2014). Mapping and Analyzing Hate Speech Online: Opportunities and Challenges for Ethiopia. Programme in Comparative Media Law and Policy, University of Oxford. Available from: http://pcmlp.socleg.ox.ac.uk/sites/pcmlp.socleg.ox.ac.uk/files/Ethiopia%20hate%20speech.pdf. [Accessed 2 Sept. 2014].*
2 *http://myanmarido.org/en. [Accessed 2 Sept. 2014].*

European Interagency Security Forum (EISF)

EISF is an independent network of Security Focal Points who currently represent 66 Europe-based humanitarian NGOs operating internationally. EISF is committed to improving the security of relief operations and staff. It aims to increase safe access by humanitarian agencies to people affected by emergencies. Key to its work is the development of research and tools which promote awareness, preparedness and good practice.

EISF was created to establish a more prominent role for security risk management in international humanitarian operations. It facilitates exchange between member organisations and other bodies such as the UN, institutional donors, academic and research institutions, the private sector, and a broad range of international NGOs. EISF's vision is to become a global reference point for applied practice and collective knowledge, and key to its work is the development of practical research for security risk management in the humanitarian sector.

EISF is an independent entity currently funded by the US Office of Foreign Disaster Assistance (OFDA), the Swiss Agency for Development and Cooperation (SDC), the Department for International Development (DFID) and member contributions.

www.eisf.eu

Editors

Raquel Vazquez Llorente and Imogen Wall.

The editors welcome comments and further submissions for future publications or the web-based project. If you are interested in contributing, please email eisf-research@eisf.eu. Imogen Wall can be contacted at imogenwall@hotmail.com.

Acknowledgments

The editors would like to thank Lisa Reilly, EISF Coordinator, for her input and advice, and especially for her comments on the initial drafts. We would also like to extend our gratitude to Tess Dury, for her research support at the initial stages of the project, Brian Shorten for sharing his expertise with us, and Crofton Black for his early guidance and, as always, his continuous support.

Suggested citation

Vazquez Llorente R. and Wall, I. (eds.) (2014) *Communications technology and humanitarian delivery: challenges and opportunities for security risk management*. European Interagency Security Forum (EISF).

Cover photo: Mary Kiperus, community health worker, uses a mobile phone for reporting to the local nurse. Leparua village, Isiolo County, Kenya. February, 2014. © Christian Aid/Elizabeth Dalziel.

Disclaimer

Other EISF Publications

If you are interested in contributing to upcoming research projects or want to suggest topics for future research please contact **eisf-research@eisf.eu**.

Briefing Papers

Security Risk Management and Religion: Faith and Secularism in Humanitarian Assistance
August 2014
Hodgson, L. et al. Edited by Vazquez, R.

Security Management and Capacity Development: International Agencies Working with Local Partners
December 2012
Singh, I. and EISF Secretariat

Gender and Security: Guidelines for Mainstreaming Gender in Security Risk Management
September 2012 – *Sp. and Fr. versions available*
Persaud, C. Edited by Zumkehr, H. J. – EISF Secretariat

Engaging Private Security Providers: A Guideline for Non-Governmental Organisations
December 2011 *Fr. version available*
Glaser, M. Supported by the EISF Secretariat (eds.)

Abduction Management
May 2010
Buth, P. Supported by the EISF Secretariat (eds.)

Crisis Management of Critical Incidents
April 2010
Buth, P. Supported by the EISF Secretariat (eds.)

The Information Management Challenge
March 2010
Ayre, R. Supported by the EISF Secretariat (eds.)

Reports

The Future of Humanitarian Security in Fragile Contexts
March 2014
Armstrong, J. Supported by the EISF Secretariat

The Cost of Security Risk Management for NGOs
February 2013
Finucane, C. Edited by Zumkehr, H. J. – EISF Secretariat

Risk Thresholds in Humanitarian Assistance
October 2010
Kingston, M. and Behn O.

Joint NGO Safety and Security Training
January 2010
Kingston, M. Supported by the EISF Training Working Group

Humanitarian Risk Initiatives: 2009 Index Report
December 2009
Finucane, C. Edited by Kingston, M.

Articles

Incident Statistics in Aid Worker Safety and Security Management: Using and Producing them
March 2012
Van Brabant, K.

Managing Aid Agency Security in an Evolving World: The Larger Challenge
December 2010
Van Brabant, K.

Whose risk is it anyway? Linking Operational Risk Thresholds and Organisational Risk Management
June 2010, (in Humanitarian Exchange 47)
Behn, O. and Kingston, M.

Risk Transfer through Hardening Mentalities?
November 2009
Behn, O. and Kingston, M.

Guides

Security Audits
September 2013 – *Sp. and Fr. versions available*
Finucane C. Edited by French, E. and Vazquez, R. (Sp. and Fr.) – EISF Secretariat

Managing The Message: Communication and Media Management in a Crisis
September 2013
Davidson, S., and French, E., EISF Secretariat (eds.)

Family First: Liaison and Support During a Crisis
February 2013 *Fr. version available*
Davidson, S. Edited by French, E. – EISF Secretariat

Office Closure
February 2013
Safer Edge. Edited by French, E. and Reilly, L. – EISF Secretariat

Forthcoming publications

Office Opening Guide