BULK DATA: POLICY IMPLICATIONS (DRAFT)

Samarajiva, Rohan;Perera-Gomez, Thavisha;

;

IDRC Grant/ Subvention du CRDI: 108008-001-Leveraging Mobile Network Big Data for Developmental Policy

Annex 14: Bulk Data: Policy Implications (DRAFT)

R. Samarajiva & T. Perera-Gomez

INTRODUCTION

In many spheres of human activity, the trend is for co-present or physical interactions and transactions to be supplemented if not replaced by virtual interactions and transactions by parties who are not co-present. For example, commerce which used to require transactions between co-present buyers and sellers, or at least their authorized agents, has been increasingly shifting to virtual space since the invention of the telegraph, leading to the current wave of massive investments in, and take up of, e commerce. The delivery of public services is increasingly being done without requiring citizens to come to government offices in the form of e government.

There is no reason why law breaking, and the deterrence, detection and punishment of such acts would be an exception. The same factors that drive the shift of lawful activities from the physical world to the virtual, such as reduction in transaction costs and ability to exert control over larger spans of time and space, apply to unlawful activities as well. In fact, law breakers have additional incentives to move away from co-present interactions and transactions. Avoidance of detection may be perceived as easier in virtual space because the agents of the law may be at a technological disadvantage vis-à-vis the law breakers. The fact that laws may be lagging criminal techniques may also make the virtual space more attractive to law breakers.

Law breakers have been using virtual space for a long time, possibly ever since the telegraph was commercially deployed. Agents of the law have also been conducting their business in virtual space for a long time. Laws have been adopted and courts have issued governing decisions. What is new and different now?

What is different now is the phenomenon popularly described as big data. “Big data” is an all-encompassing term for any collection of data that is very large or complex, and therefore difficult to analyze using conventional data processing applications. Big data was always there, but now the conditions exist for low-cost, quick analysis (Mayer-Schonberger & Cukier, 2013). The focus of the present discussion is on the subset of big data known as transaction-generated data (also described as “data exhaust”) arising from the day-to-day behaviors of persons and the technological devices closely associated with them.

Surveillance is used in a non-pejorative sense in this report. Its dictionary meaning is to keep a close watch over someone or something. In social theory it is described as the control of information and the superintendence of the activities of some groups by others (Giddens, 1987, p. 2). The state is a “human community that (successfully) claims the monopoly of the legitimate use of physical force within a given territory” (Weber, 1918). Ensuring law and order is therefore a core function of the state. To maintain that monopoly it is necessary to superintend the activities of groups that challenge the state. Every state engages in surveillance.

States have always sought to control groups within their jurisdiction. At different times and under different state forms, this control has been exercised through different technologies of surveillance, among other measures. In no case has control been total. Surveillance has always been tempered by practical limitations and by constraints imposed to safeguard the rights of subjects. In modern times, state surveillance even for the control of subversion has been subject to constraints associated with civil liberties.

The balance between the requirements of the state and individuals with regard to behavioral big data is being set at the present time through practice and policy discussion. It is described differently in different countries. In this report the term "bulk surveillance" is used to describe the obtaining and analysis of behavioral big data relevant to maintenance of law and order, broadly defined.

Two main types of data are captured: contents of the communication as well communications data or metadata, that is, information about the communication (who, what, when and where). For instance, when considering mobile network data, whenever a telecom service is being used, a record is generated for billing and record purposes. Call Detail Records (CDRs) capture certain types of transactions (such as when a call was made, the recipient and duration of the call as well as the cell tower the call was connected to). Internet connection records are also included within the broad definition of CDRs. La Rue (2013, p. 3) defines communication data as:

> Information about an individual's communications (e-mails, phone calls and text messages sent and received, social networking messages and posts), identity, network accounts, addresses, websites visited, books and other materials read, watched or listened to, searches conducted, resources used, interactions (origins and destinations of communications, people interacted with, friends, family, acquaintances), and times and locations of an individual, including proximity to others.

The acquisition of data in large volumes–bulk data[1]–could include both the communication itself as well as the contents of communication. Bulk data reportedly allow authorities to obtain information not readily accessible using conventional targeted means, for instance identifying behavior patterns and networks (*Operational Case for Bulk Powers*, n.d.)

Laws in certain countries mandate the retention of various types of communications data of users– this could span from 12 months in the UK (BBC, 2016), two years in Australia (BBC, 2015), to five years in South Africa (Privacy International, 2017) and Brazil (Privacy International, 2017).[2] As expressed by a Danish government representative:[3] "It's impossible to know beforehand which data might be relevant in the future for solving criminal cases. Data retention provides investigators with the benefit of hindsight…." (Bowcott, 2016). In many countries, the rules affecting retention of bulk data and who is allowed access to them are not public.

In India, the government is developing a Centralized Monitoring System (CMS) to centralize the interception of communication data, bypassing service providers.[4] Moreover, it is reported that the Network Traffic Analysis software is intended to intercept web traffic, filtering it to detect words such as 'kill' and 'bomb' (Privacy International, 2017).

---

[1] Based on the Operational Case for Bulk Powers. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/504187/Operational_Case_for_Bulk_Powers.pdf

[2] UK – communications data; Australia – telecommunications metadata; South Africa – telecommunications metadata; Brazil – data from landline service providers and mobile service providers.

[3] As reported by the Guardian, April 12 2016: https://www.theguardian.com/world/2016/apr/12/mp-david-davis-calls-limit-uk-surveillance-powers-european-court-justice

[4] For more information, see 2016-17 annual report of the Department of Telecommunications, http://www.dot.gov.in/sites/default/files/Annual%20Report%202016-17.pdf?download=1

Proponents claim that the analysis of bulk data can help authorities detect emerging threats, particularly in situations where they do not have enough information to conduct targeted surveillance. Moreover, after a person of interest (POI) has been identified, an analysis of the communication patterns of individuals the POI has communicated with would help identify a broader network of accomplices. Proponents state that in fast paced investigations, bulk data could provide a sense of directionality enabling prioritization of resources to handle the more serious threats. Once bulk surveillance is used to identify a POI, targeted surveillance techniques could then be deployed to obtain more information on the target.

In many cases laws governing surveillance have not kept pace with technological advancements. For instance, consider the constitutional challenge launched against South Africa's Regulation of Interception of Communications and Provision of Communication-Related Information Act (RICA) by AmaBhungane Centre for Investigative Journalism in 2017—one of the areas contested being the lack of regulation on bulk interception.

MODALITIES: TARGETED VS. BULK SURVEILLANCE

Targeted surveillance is typically conducted on an identified person or group that the state deems to be reasonably under suspicion. Law enforcement agencies have leveraged targeted surveillance to monitor suspected terrorists and criminals (Council of Europe, 2015). Few oppose this type of surveillance.

The other form is bulk surveillance, which captures data across a broad segment of population to identify potential persons of interest. Thus, bulk surveillance may be seen as a precursor of targeted surveillance.

Instead of the contested term 'mass' surveillance (Cannataci, 2017), Anderson (2016) focuses on the use of bulk powers such as bulk interception, bulk acquisition and bulk equipment interference. Thus the term 'bulk surveillance' is used in this report.

In the use of surveillance to combat crime and terrorism, it has been argued that focusing efforts on the small percentage engaged in unlawful acts would enable law enforcement and intelligence agencies to be more efficient (Abraham, 2014). Moreover, this would ensure that there is no indiscriminate monitoring of a population. As Abraham (2014) further states, "In post-facto surveillance, those people who were or are in some way connected to an event are targeted." Thus, permission to violate an individual's privacy would only be sought after a probable cause has been established and there is 'reasonable' suspicion (Song, 2014). This approach is feasible only when the relevant persons of interest have already been identified.

Proponents of the above approach differentiate between physical space and virtual space, believing that all suspects would be identified and probable cause established in the former with targeted surveillance then being conducted in virtual space. This does not correspond to reality. The solution depends on the assumption that all potential law breakers would fall within the defined subset identified for targeted surveillance with none being included in the excluded-from-surveillance subset. This is patently unrealistic. Law breakers have every incentive in the world to get themselves included in the excluded subset. Some transgressions of the law occur completely in virtual space. Their investigation have necessarily to take place fully in virtual space.

INSTANCES OF BULK SURVEILLANCE

Among other things, the analysis of bulk data can help identify POI and provide insights on the POI, including their social networks, behavior and movement patterns. It can also be used to provide more context on the network environment, and guide decisions on resource prioritization (Anderson, 2016). It is possible to classify the use of bulk data to identify persons of interest and potential targets (places) before an act of terror is committed, and to identify perpetrators after an act of terror. This can be illustrated as follows:

Table 1: Forms of Bulk Surveillance

	1. Person(s)	2. Place
1. Pre-event	**1.1**	**1.2**
2. Post-event	**2.1**	**2.2**

Source: Authors, 2017

3.1 Pre-event Surveillance

Everyone in counter-terrorism has a deep interest in preventing terrorist acts, not in simply catching the terrorists after the fact, and this is where the appeal of 'mass'/bulk surveillance lies. From the citizen's point of view, prevention even of petty crime would be preferred over detention, after the fact. Terrorist atrocities in Pakistan, Afghanistan, Iraq, and Europe in the recent past have strengthened the case for expansion of surveillance powers to assist in capturing terrorists. Macaskill & Dance (2013) reported that the US National Security Agency's (NSA) justification of data collection is: "The NSA say it needs all this data to help prevent another terrorist attack like 9/11. In order to find the needle in the haystack, they argue, they need access to the whole haystack."

In addition to terrorism, there is opportunity to leverage such forms of surveillance to combat crime, helping to identifying hotspots for potential crime. The use of 'bulk surveillance' can highlight anomalies that would not be visible otherwise.

3.1.1 Pre-event – Person(s) [1.1]

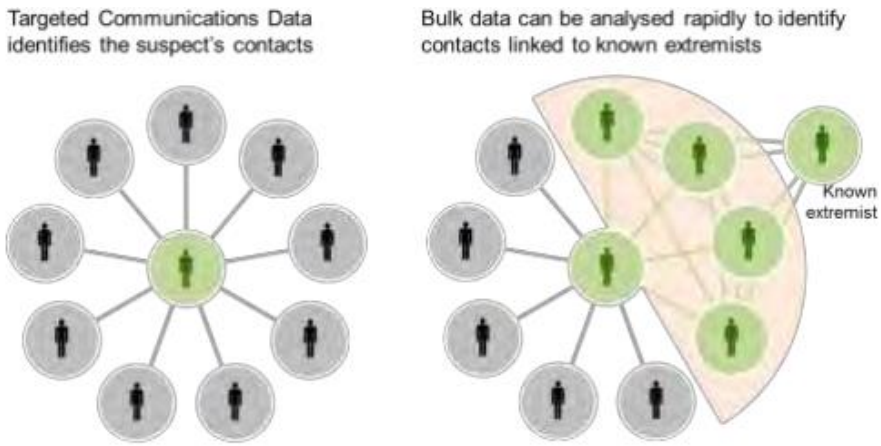
There have been few concrete/public examples of the use of bulk surveillance in identifying persons of interest before a crime/act of terrorism has been used. In the wake of Edward Snowden's revelations, authorities have claimed that that internet surveillance played a key role in averting a plot to 'bomb the New York city subway' in 2009 by Najibullah Zazi (Pilkington & Watt, 2013). However, the role of bulk surveillance in this has been questioned (Bergen, Sterman, Schneider, & Cahall, 2014; Pilkington & Watt, 2013).

In 2016, it was reported that Fresno, a city in California was using a tool that generated "threat scores" for individuals, (Sadowski, 2016), analyzing "billions of data points, including arrest reports, property records, commercial databases, deep web searches and the man's social- media postings" (Jouvenal, 2016). While the software scans an individual's names against public data to generate a color-coded score, how the score is tallied is unclear, neither the police nor the public having access to this (Jouvenal, 2016).

Anderson (2016) provides an example of UK intelligence service, MI5's use of bulk personal data sets[5] to identify persons of interest who were likely to pose a threat to the 2012 London Olympics. The analysis of this data enabled the agency to sift through persons of interest, ruling out some, allowing them to focus more on those with greatest value. Moreover, terrorists often operate as a part of a larger network and bulk data can be used to quickly identify connections:

> In 2010, a network of terrorists – comprising groups in Cardiff, London and Stoke-on-Trent - planned a series of bomb attacks at several symbolic locations in the UK, including the London Stock Exchange. Complex analysis of bulk acquisition data[6] played a key role in identifying the network. The task was made particularly challenging by the geographical separation of the groups. Nine members of the network were subsequently charged and pleaded guilty to terrorism offences relating to the plot. Eight members of the network pleaded guilty to engaging in conduct in preparation for acts of terrorism (Anderson, 2016, p. 174).

Figure 1: Uncovering networks using bulk communications data



Source: Operational Case for Bulk Powers, n.d. [7]

Once a POI has been identified, the individuals communications data can be used identify individuals the POI has interacted with. The common practice is to extend the analysis to contacts of contacts, as shown in Figure 1. Analysis of bulk data can help identify which of these contacts had links to extremists already on the radar of intelligence agencies.

3.1.2 Pre-event – Place [1.2]

[5] According to MI5, Bulk Personal Datasets refer to: "sets of personal information about a large number of individuals, the majority of whom will not be of any interest to MI5. https://www.mi5.gov.uk/bulk-data

[6] According to MI5, bulk communications data acquisition is: "who", "where", "when", "how" and "with whom" of communications, but not what was written or said. https://www.mi5.gov.uk/bulk-data

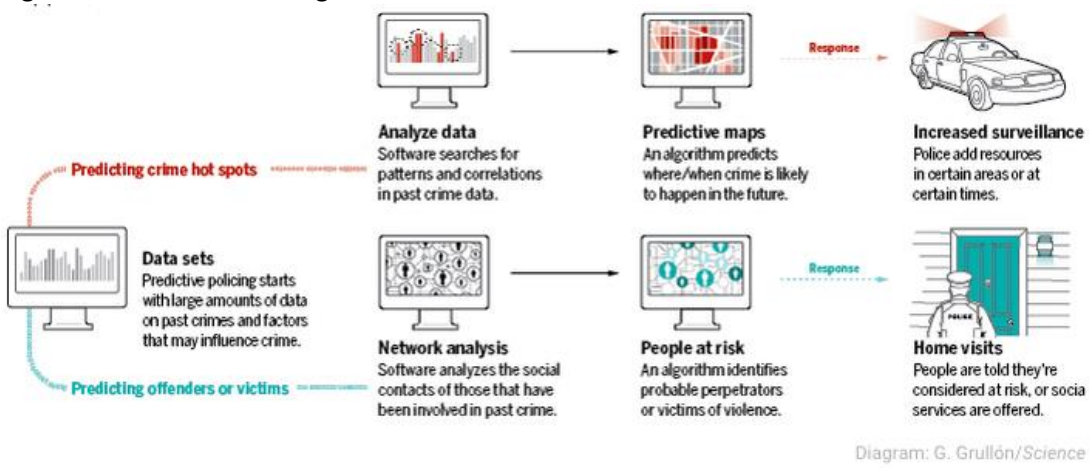
[7] For more information, see: https://www.gov.uk/government/publications/investigatory-powers-bill-overarching-documents

An alternative to identifying a person is the adoption of a place-centric approach. This means identifying a potential place for a crime/act of terrorism before it occurs.

Bogomolov, et al. (2014) sought to identify crime hotspots by leveraging mobility data derived from mobile phones, as well as open data such as sales of residential property, weather data, transportation data, data on past crimes, etc. When the "experimental results" were compared against real crime data, it showed an accuracy of nearly 40% in predicting whether or not a particular location would be a crime spot.

Similarly, in predictive policing, algorithms are used to forecast places where crime is likely to occur (Shapiro, 2017). Companies such as Predpol offer law enforcement agencies solutions that leverage historical data such as type of crime, location, date and time to predict crime (place and time), not predict the perpetrator. By leveraging just these variables, PredPol argues that the risk of 'discriminatory profiling' is minimized. In September 2017, the New India Express reported that a big data solution was being developed for law enforcement agencies in five states in India to predict crime (Sharma, 2017). The solution would model those deployed in the US and take into account crime data as well as other data such as weather, dates (e.g., pay day).

Figure 2: Predictive Policing: How it works



Source: Hvistendahl (2016)

3.2 Post Event Surveillance

Even in a post-event surveillance scenario, the identification of persons of interest itself may entail casting a wider net. For instance, if a crime was committed at a particular place, at a particular time, identifying possible suspects may mean first identifying all the people who were in the vicinity at that particular time (for example, using a cell tower dump). The analysis of call detail records can provide important insights into the mobility patterns and social networks of individuals, and provide vital information that could show important connections that solve crimes by helping to show what suspects were doing/where suspects were before, during and after a crime was committed.

3. 2.1 Post-event –Person [2.1]

Moreover, communications data itself maybe leveraged to develop a profile of the individual, associations, social network etc. For instance:

> Following a failed terrorist attack in London in 2007, the security and intelligence agencies were able to confirm that the perpetrators were the same as a group who had carried out another attack shortly afterwards. This was achieved in a matter of hours through the analysis of bulk communications data, and was vital in understanding the scale of the threat posed in a fast-moving post-incident investigation, because of the ability to identify connections at speed; it would not have been possible to do this at speed by relying on requests for targeted communications data. Through further analysis of communications data, the investigation went on to identify people who had had extensive contact with telephones used in the London attack. This enabled the security and intelligence agencies and police to establish at speed, that no further attacks were planned. The operation led to a successful prosecution. (Operational Case for Bulk Powers, 2016, p. 41).

Solutions such as those offered by Palantir to Law Enforcement Agencies enable use of a single portal to conduct searches across numerous data systems– helping authorities to identify relationships between seemingly unrelated sources. For instance in California, the Palantir system has integrated the California Law Enforcement Telecommunications System as well as data from automated license plate readers, and other databases (Harris, 2017).

3. 2.1 Post-event –Place [2.2]

This was not explored on the premise that once an event has occurred, there is little point looking for the place.

SELECTED ISSUES

Bulk surveillance has been criticized on numerous grounds including, but not limited to intrusion of privacy of individuals and suppression of freedom of expression.

Suppression of civil liberties

Samarajiva and Lokanathan (2016) state that techniques used to enable location-based services can also be used to track movements of groups or individuals for other purposes. Indications exist that mobile network big data are already being used by governments to identify and control gatherings. In January 2014 text messages warned protestors in Kiev, Ukraine, that they were participants in a mass riot: "Dear subscriber, you are registered as a participant in a mass riot." The mobile operators MTS and Kyivsta issued statements claiming they were not responsible for the messages. Privacy International (n.d.) states that there is “a strong suggestion” that this was done using IMSI catchers.

Data insecurity

As the volume and value of aggregated data increases, the harms that can be caused by the data falling into wrong hands or being distorted increase. Increasingly frequent reports of breaches of security leading to troves of big data including sensitive personally identifiable information falling into the hands of data thieves indicates that the security of bulk data in the custody of government or its agents a serious problem.

Opacity of algorithms

A heavy reliance on algorithms to generate insights for law enforcement agencies have also raised concern, particularly when such algorithms are developed by third parties, who do not share the methodology with law enforcement agencies and with the public, in essence

creating a black box. This is further exacerbated by the fact that if the input data were biased, then the results generated by the algorithm would reflect these biases. For instance, if an algorithm were only provided crime data for black people, then black neighborhoods would be flagged for police patrol (Patel, 2015).

Panoptical Effect

Moreover, another concern is that when an individual has reason to believe that he may be being watched, there is a tendency to modify behavior, as people strive to conform (Panoptical effect). For example, certain behaviors will be avoided because of the probability of observation.

Penney (2016) has demonstrated that concerns about surveillance affect individuals' online activities, adversely impacting the very purpose of surveillance. For instance, it is in the best interest of lawbreakers to pass of as law-abiding citizens, striving for conformity. Thus, if they are aware that their mobile phones are being monitored, they will strive to work around it. For instance, avoiding the use of mobile phones. However, the use of avoidance measures may themselves be an indicator.

Legislation and court rulings on bulk data

In 2016, the UK passed the controversial Investigatory Powers Act. Among other provisions, it requires internet and communications companies to collect internet browsing history of their customers for up to a year, and it allows intelligence organizations MI5 and Government Communications Headquarters (GCHQ) to conduct bulk communications data acquisition (McGoogan, 2017).

However, in recent years, there have been efforts to address concerns around bulk retention:

"In the Digital Rights Ireland judgment of 2014,[8] the Court of Justice declared invalid the directive on the retention of data[9] on the ground that the interference, by the general obligation to retain traffic data and location data imposed by that directive, in the fundamental rights to respect for privacy and the protection of personal data was not limited to what was strictly necessary." (Court of Justice of the European Union, 2016, p.1)

In 2016, the European Court of Justice ruled against the indiscriminate retention of data. According to the Press release of the Court of Justice of the European Union (2016, p.1), "EU law precludes a general and indiscriminate retention of traffic data and location data, but it is open to Members States to make provision, as a preventive measure, for targeted retention of that data solely for the purpose of fighting serious crime, provided that such retention is, with respect to the categories of data to be retained, the means of communication affected, the persons concerned and the chosen duration of retention, limited to what is strictly necessary. Access of the national authorities to the retained data

---

[8] Joined Cases: C-293/12 and C-594/12 Digital Rights Ireland and Seitlinger and Others, see Press Release No 54/14. For more information: http://curia.europa.eu/juris/document/document_print.jsf?doclang=EN&text=&pageIndex=0&part=1&mode=lst&docid=150642&occ=first&dir=&cid=314051

[9] Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC (OJ 2006 L 105, p. 54).

must be subject to conditions, including prior review by an independent authority and the data being retained within the EU."

In September 2017, "the Investigatory Powers Tribunal (IPT) ruled that the European court of Justice (ECJ) should decide whether the UK's bulk collection of communications data, tracking personal use of the web, email, texts and calls, is legal," in the judgment of a case brought to IPT of the UK by Privacy International against MI5, MI6 and GCHQ (Travis, 2017).

Even in the United States, there is interest in strengthening digital privacy protection. A Supreme Court decision in 2014 (Riley vs. California), the Court ruled that police could not search cellphones without a warrant (Liptak, 2014). The New York Times (Liptak, 2017) reports that the robbery of a Radio Shack store could potentially impact the Fourth Amendment's protections on privacy. In the case Carpenter v. United States, No. 16-402 – with a decision expected in June 2018 – the Supreme Court is considering if the Fourth Amendment was violated when prosecutors collected location data of a prime suspect's phone from cellphone companies. The companies provided records for 127 days. It is expected that the reasoning of the Supreme Court would also be applied to other sources: bank records, internet search data, email messages, etc.

The New York Times (Liptak, 2017) quotes Chief Justice John G. Roberts:

> "Modern cellphones are not just another technological convenience. Even the word cellphone is a misnomer. They could just as easily be called cameras, video players, Rolodexes, calendars, tape recorders, libraries, diaries, albums, televisions, maps or newspapers."

Issues of Oversight

As part of efforts to automate the lawful interception process, the Indian government has been rolling out a Centralized Monitoring System (CMS) that is intended to take away the process of manual intervention by telecom operators in obtaining electronic data on targets through a secure connection. In addition to intercepting mobile phone communication, this would also cover landlines as well as the Internet. Telecommunications Minister, Ravi Shankar Prasad, stated that the CMS would enable the "secure flow of intercepted communication on near real time basis between law enforcement agency and Telecom Service Providers on secured and dedicated CMS network" (Times of India, 2016). The Center for Internet & Society has raised concerns that this would enable governments to bypass mobile network operators—who are typically informed of interception requests—and gain direct access to the data, with interception requests now being fielded by the CMS authority (Xynou, 2014).

Lawful Access and Transparency

While the laws compelling communications service providers to provide governments with access to communication data of their customers vary by country, they are typically for the purposes such as public safety, national security, terrorism and crime. Communications service providers receive various types of legal demands: for telecom operators in the US, these can range from subpoenas, court orders and emergency requests to search warrants and national security letters as well as foreign intelligence surveillance act orders (T-Mobile, 2015).

The multi-country telecommunications operator, Telenor (2015) segments government requests into five categories, namely,

- Communications data – (i.e. retained/historical data)
- Lawful interception (real-time interception of communication)
- Network shutdowns (partial or complete shutdown of an operator network)

Content restrictions (restrictions that are imposed on the distribution of electronic content), and

- Content distribution (distribution of information from the government to mobile users).

However, not all forms of lawful access maybe disclosed to the public (for example, laws may inhibit this or governments may request that this not be shared). Moreover, in addition to lawful access, agencies/authorities may conduct surveillance without the cooperation of mobile network operators.

Communications service providers are faced with balancing their customers' interests related to privacy and freedom of expression whilst adhering to the local laws of the countries they operate in. They try to address this by the publication of reports, typically called transparency reports, in which they reveal, where possible, government requests to obtain customer data. Facebook, Twitter, Apple, Google, Airbnb are among the technology and communications service companies who share this information.

Even in this case however, telecom operators in some countries are not legally allowed to disclose statistics on interception requests. Moreover, in select countries, government may have direct access to operator data. Similarly, even if disclosure laws are ambiguous, governments may request authorities to not share statistics (Telenor, 2016).

While there appears to be a trend of telecom operators promoting greater transparency, some telecom operators appear to be believe that it should ideally be governments that disclose this information and not the operators. For one, no single operator would have clarity on the number of requests made in the country as they have visibility to their network. One solution is to encourage all operators in the country to disclose relevant information.

However, there may be differences in reporting. For instance, while one operator may disclose the number of requests made, another may disclose either a mix of total statistics on the number of accounts, devices, subscribers or services that were targeted (Vodafone, 2014). There is a difference here. A request may target from one to multiple accounts. Additionally, there may be multiple requests to access different types of data for one customer. Furthermore, a government may issue the same request to all operators and if each operator reports it any of the above ways, this would contribute to the lack of clarity in the space. Telecom operators argue that Governments are better positioned to disclose requests made to all operators in the country.

Conversely, if communication data is being used to fight crime, terrorism, strengthen national security and improve public safety, it is in the interest of law enforcement agencies to keep the public remain unaware or unsure of the nature and scope of surveillance being conducted. The rationale would be that it would make the taking of evasive measures more difficult for law breakers.

CONCLUSION

It is useful to approach problems of bulk data and surveillance based on the understanding that the physical and virtual are best seen as a continuum with different qualities

predominating in different parts. Virtual space is not radically different from physical space nor is it an appendage of physical space. This would require solutions to problems of law-breaking, deterrence, detection and punishment being derived from common principles and, wherever possible, from laws that are not "medium-specific," by which is meant that what is a transgression should not change depending on whether it occurred in physical or virtual space. The forms of investigation, evidence gathering, and so on would have to be different, but even here, it is advisable to anchor the practices on principles that are common to the both physical and virtual space.

The common-law tradition where judges devise remedies for problems brought before them works in that manner. For example, US judges who were asked to decide on the legality of supply of pornographic material using postal or electronic means, extended the principle community standards originally devised for co-present transactions to those that occurred in virtual space. The exception is where legislatures come up new laws that are specific to technologies.

Bulk surveillance can take place before an infraction occurs (pre-event surveillance) and after (post-event surveillance). Pre-event bulk surveillance can focus on people or on places. In the case of post-event surveillance, the place is already known. Therefore it applies only to people.

Post-event bulk surveillance is the least problematic of the three forms. One could say that it is the virtual equivalent of investigating all who were present in a location where a crime occurred. The claim, made by some, that bulk surveillance should not be conducted on any other than a subset of persons against whom probable cause has been established is not consistent with long-established investigatory practices in physical space. Critics may point to the practice of investigating not just the contacts of the persons present at the time of the event but also the contacts of the contacts as something that differentiates the virtual from the physical. It is true that this can be done more rigorously in the virtual context than in the physical. However, there is nothing new in the practice itself.

The questions then have to be limited to harms that may be caused by the analysis of post-event data obtained through bulk surveillance. One obvious harm is from secondary uses disconnected from the original investigation.

The Nepal Supreme Court decision that found illegal the retention and use for different purposes of mobile network bulk data collected to investigate a murder points to a solution. Laws and investigatory practices will have to be set in place to ensure than bulk data collected for a specific investigation will have to be ring-fenced and access to the data in raw or processed form limited. The model would be no different from that which applies to paper-based investigatory records and tangible evidence. These items are kept in controlled environments with recorded maintained of authorized access. In the case of bulk data, the safeguards and logs would have to be designed appropriately for the medium they would be stored in.

For how long should the data be retained? Under what conditions should bulk data collected for one investigation be used for another investigation after the fact, or for the more problematic purpose of pre-event investigation intended to deter or prevent unlawful acts?

Bulk surveillance that is conducted prior to an event either to identify or safeguard places where laws are likely to be broken or to identify or deter persons likely to engage in unlawful acts is where much of the controversy is focused on. Because humans know that predicting the future is difficult and is prone to error, there is a general suspicion about government engaging in such actions, especially those that may infringe civil liberties.

One may engage in abstract discussions about the propriety of agents of the state seeking to prevent or deter crime. But in the concrete circumstances of societies facing criminal or terroristic threats, there would be no debate about the value of deterring or preventing unlawful acts.

There is a concern that unjust laws may be enforced through analysis of data obtained through bulk surveillance. For example, there is a difference of opinion in most societies about political protests seeking to topple the government in power. Some would see any action to identify protestors or deter them as wrong and would oppose the use of bulk surveillance and data analytics for such purposes. Others would point to existing laws against unauthorized assembly and would see nothing wrong in using normal investigatory techniques including bulk surveillance to enforce the law. The question of what laws are just and which are unjust cannot be resolved in this paper.

The questions then have to be limited to harms that may be caused by attempting to predict through the analysis of data obtained through bulk surveillance. One obvious harm is from secondary uses disconnected from the original investigation. The issues of retention of and who has access to the data are no different from those discussed above, in relation to post-event bulk surveillance.

Using bulk data to predict may cause harm because the predictions are erroneous. An example is analysis that points to a particular individual as being likely to engage in unlawful activity. This is may lead to actions that harm an individual's civil liberties or on the operations of an organization. The safeguards against this form of harm are currently being discussed in the broader context of algorithmic fairness (e.g., Bornstein, 2017).

What of the impacts on civil liberties of an individual or on the operations of an organization when the prediction is correct? In one scenario, the prediction will deter. If the prospective law breaker understands that the plot has been discovered, he/she may abandon it. In some cases such the most heinous crimes, it common to define the attempt also as a crime (e.g., attempted murder).

In the second scenario, the prospective law breaker continues his/her activities unaware that the plot has been discovered. Given the difficulties of proving the attempt if the process is aborted too early, sometimes investigators may allow the activity to continue up to a point when they consider enough evidence has been amassed to prosecute.

Both scenarios may occur in the physical world as well as in the virtual. The only thing that may be unique to bulk surveillance has to do with disclosure. It is relatively simple to understand why investigators would not want a suspect to know that his/her plans have been discovered. In the case of searches of physical spaces, most countries require warrants, which require disclosure. In the case of bulk as well as individual searches in virtual space the search is not obviously visible as in physical space. Therefore, it requires an additional decision on whether or not to disclose a search.

The common practice with regard to post-event investigations is not to disclose to suspects that their communications are being monitored under warrant. What should the principle be with regard to bulk surveillance? The much criticized practice under the Foreign Intelligence Surveillance Act (FISA) by the courts established under it was to prohibit all disclosure. An acceptable middle ground may be to allow the companies subject to bulk surveillance orders to report the broad outlines of the activity as Telenor does (see above). Having an independent authority such as a court consider each request for bulk surveillance and authorize each is necessary to prevent abuse. This is not the case in many countries. Disclosing the level of bulk surveillance without being specific would be an additional safeguard against abuse.

REFERENCES

Abraham, S. (2014). Comment to The 80:20 rule in electronic communication surveillance [Comment on Blog post]. Retrieved from http://lirneasia.net/2014/03/the-8020-rule-in-electronic-communication-surveillance/

Anderson, D. (2016). Report of the Bulk Powers Review. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/546925/56730_Cm9326_WEB.PDF

BBC News (2015, October 13). Australia begins mass data retention under new law. Retrieved from http://www.bbc.com/news/world-australia-34513124

BBC News (2016, December 21). EU data retention ruling goes against UK government, Retrieved from http://www.bbc.com/news/uk-politics-38390150

Bergen, P., Sterman, D., Schneider, E., & Cahall, B. (2014). Do NSA's Bulk Surveillance Programs Stop Terrorists? Washington, D.C. Retrieved from https://static.newamerica.org/attachments/1311-do-nsas-bulk-surveillance-programs-stop-terrorists/IS_NSA_surveillance.pdf

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014). Once upon a crime: towards crime prediction from demographics and mobile data. In 16th international conference on multimodal interaction, ACM (pp. 427–434).

Bornstein, A. M. (2017, December 21). Are Algorithms Building the New Infrastructure of Racism? Nautilus. Retrieved from http://nautil.us/issue/55/trust/are-algorithms-building-the-new-infrastructure-of-racism

Bowcott, O. (2016, April 12). MP calls for limit on UK surveillance powers as EU test case opens. The Guardian. Retrieved from https://www.theguardian.com/world/2016/apr/12/mp-david-davis-calls-limit-uk-surveillance-powers-european-court-justice

Cannataci, J. (2017). Report of the Special Rapporteur on the right to privacy, Joseph A. Cannataci.

Court of Justice of the European Union, P. R. N. 145/16. (2016, December 21). The Member States may not impose a general obligation to retain data on providers of electronic communications services. Retrieved from https://curia.europa.eu/jcms/upload/docs/application/pdf/2016-12/cp160145en.pdf

Giddens, A. (1987). The nation state and violence. Berkeley CA: University of California Press.

Harris, M. (2017, August 09). How Peter Thiel's Secretive Data Company Pushed Into Policing. Wired. Retrieved from https://www.wired.com/story/how-peter-thiels-secretive-data-company-pushed-into-policing/

Hvistendahl, M. (2016, September 28). Can 'predictive policing' prevent crime before it happens? Science. Retrieved from http://www.sciencemag.org/news/2016/09/can-predictive-policing-prevent-crime-it-happens

Jouvenal, J. (2016, January 01). The new way police are surveilling you: Calculating your threat 'score'. The New Washington Post. Retrieved from https://www.washingtonpost.com/local/public-safety/the-new-way-police-are-surveilling-you-calculating-your-threat-score/2016/01/10/e42bccac-8e15-11e5-baf4-bdf37355da0c_story.html?utm_term=.1f7f6d1b827e

La Rue, F. (2013). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression , Frank La Rue. United Nations Human Rights Council. Retrieved from https://www.google.lk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjBsPCC08XWAhUVSo8KHVASAyUQFggkMAA&url=http%3A%2F%2Fwww.ohchr.org%2FDocuments%2FHRBodies%2FHRCouncil%2FRegularSession%2FSession23%2FA.HRC.23.40_EN.pdf&usg=AFQjCNFiNEUL0cux7EhiuCa9J0145O1M-Q

Liptak, A. (2017, November 27). How a Radio Shack Robbery Could Spur a New Era in Digital Privacy. The New York Times. Retrieved December 05, 2017, from https://www.nytimes.com/2017/11/27/us/politics/supreme-court-fourth-amendment-privacy-cellphones.html?emc=eta1

Liptak, A. (2014, June 2014). Major Ruling Shields Privacy of Cellphones. The New York Times. Retrieved from https://www.nytimes.com/2014/06/26/us/supreme-court-cellphones-search-privacy.html?_r=0

Macaskill, E., & Dance, G. (2013, November 01). NSA Files:Decoded - what the revelations mean for you. The Guardian. Retrieved from https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded#section/1

Mayer-Schonberger, V. & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. London: John Murray

McGoogan, C. (2017, June 30). Government's surveillance powers to be challenged in High Court. The Telegraph. Retrieved from http://www.telegraph.co.uk/technology/2017/06/30/governments-surveillance-powers-challenged-high-court/

Nakashima, E. (2017, August 14). Tech firm is fighting a federal demand for data on visitors to an anti-Trump website. The New Washington Post. Retrieved from https://www.washingtonpost.com/world/national-security/tech-company-is-fighting-a-federal-order-for-ip-addresses-to-find-visitors-to-an-anti-trump-website/2017/08/14/a65b7544-8152-11e7-b359-15a3617c767b_story.html?utm_term=.59674a681140

Operational Case for Bulk Powers (n.d.).Gov.uk. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/504187/Operational_Case_for_Bulk_Powers.pdf

Council of Europe, Parliamentary Assembly, (2015). Mass Surveillance. Retrieved from http://www.assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-EN.asp?fileid=21692&lang=en

Patel, F. (2015, December 03). Be Cautious About Data-Driven Policing. The New York Times. Retrieved from https://www.nytimes.com/roomfordebate/2015/11/18/can-predictive-policing-be-ethical-and-effective/be-cautious-about-data-driven-policing

Penney, J. (2016). Chilling Effects: Online Surveillance and Wikipedia Use . Berkeley Technology Law Journal, Vol. 31, No. 1, p. 117, 2016. Available at SSRN: https://ssrn.com/abstract=2769645

Pilkington, E., & Watt, N. (2013, June 12). NSA surveillance played little role in foiling terror plots, experts say. The Guardian. Retrieved from https://www.theguardian.com/world/2013/jun/12/nsa-surveillance-data-terror-attack

Privacy International. Communications Surveillance. (n.d.). Retrieved from https://www.privacyinternational.org/node/10
Privacy International. Mass Surveillance (n.d.). Retrieved from https://www.privacyinternational.org/node/52
Privacy International. State of Privacy Brazil. (2017, July 03). Retrieved from https://www.privacyinternational.org/node/979
Privacy International. State of Privacy India. (2017, July 26). Privacy international.org. Retrieved from https://www.privacyinternational.org/node/975
Privacy International. State of Privacy South Africa (2017, June 28). Retrieved from https://www.privacyinternational.org/node/968
Sadowski, J. (2016, February 4). Police data could be labelling 'suspects' for crimes they have not committed . The Guardian. Retrieved from https://www.theguardian.com/technology/2016/feb/04/us-police-data-analytics-smart-cities-crime-likelihood-fresno-chicago-heat-list
Samarajiva, R. (2011, May). No Internet and telephone in a million dollar house [Blog post]. Retrieved from http://lirneasia.net/2011/05/no-internet-and-telephone-in-a-million-dollar-house/
Samarajiva, R. (2014, March 18). The 80:20 rule in electronic communication surveillance [Blog post]. Retrieved from http://lirneasia.net/2014/03/the-8020-rule-in-electronic-communication-surveillance/
Samarajiva, R., & Lokanathan, S. (2013). Using behavioral big data for public purposes: Exploring frontier issues of an emerging policy arena.
Shapiro, A. (2017). Reform predictive policing. Nature, 541(7638), 458-460. Retrieved from http://www.nature.com/news/reform-predictive-policing-1.21338
Song, S. (2014). Comments to The 80:20 rule in electronic communication surveillance [Comment on Blog post]. Retrieved from http://lirneasia.net/2014/03/the-8020-rule-in-electronic-communication-surveillance/
Sharma, V. (2017, September 23). Indian police to be armed with big data software to predict crime. The New Indian Express. Retrieved from http://www.newindianexpress.com/nation/2017/sep/23/indian-police-to-be-armed-with-big-data-software-to-predict-crime-1661708.html
T-Mobile. (2015). T-Mobile Transparency report for 2015. Retrieved from https://newsroom.t-mobile.com/content/1020/files/2015TransparencyReport.pdf
Telenor (2015). Authority Requests for Access to Electronic Communication – Legal Overview. Retrieved September 27, 2017 from https://www.telenor.com/wp-content/uploads/2015/05/GOVERNMENT-ACCESS-REPORT_05.pdf
Telenor Group. (2016, April). Authority Requests Disclosure report. Retrieved from https://www.telenor.com/wp-content/uploads/2015/05/Authority-Requests-Disclosure-Report-2015_04.pdf
The Times of India (2016, May 04). Govt setting up CMS for lawful interception: Prasad. Retrieved from http://timesofindia.indiatimes.com/city/delhi/Govt-setting-up-CMS-for-lawful-interception-Prasad/articleshow/52110186.cms
Travis, A. (2017, September 08). Tribunal says EU judges should rule on legality of UK surveillance powers . The Guardian. Retrieved from https://www.theguardian.com/world/2017/sep/08/snoopers-charter-tribunal-eu-judges-mass-data-surveillance
Vodafone (2014). Sustainability Report. Retrieved from https://www.vodafone.com/content/dam/sustainability/2014/pdf/vodafone_full_report_2014.pdf

Xynou, M. (2014, January 30). India's Central Monitoring System (CMS): Something to Worry About? Retrieved from https://cis-india.org/internet-governance/blog/india-central-monitoring-system-something-to-worry-about