



SOCIAL MEDIA SURVEILLANCE AND STATE CONTROL: A CROSS-COUNTRY ANALYSIS

Sheikh Adnan Ahmed Usmani

Ph.D. Scholar, Dept. of Computer Science,

Federal Urdu University of Arts Science and Technology, Karachi

Email: a2usmani@gmail.com

Abstract

Social media has emerged as a space for political expression and also a significant monitoring and control area by the state. This paper examines social media surveillance through a computer science lens, considering it a multi-layered technical system that includes traffic control at the network level, identity and content requirements at the platform level, AI-powered analytics, and enforcement systems. The paper is based on empirical data from the 2022 Freedom on the Net (FOTN) dataset to build a cross-country comparative framework that spans eight countries with significantly different scores on internet freedom, ranging from the United States to China, Iran, Russia, Pakistan, Turkey, India, Brazil, and others. The analysis visualizes the known surveillance infrastructure of each country on a four-part technical map, which breaks down into network infrastructure, platform and identity, analytics, and AI, and finally enforcement. The paper shows that the level of technological sophistication of a state's surveillance stack, in terms of its ability to conduct deep packet inspection, to link people to real names or biometric data, and to classify contents automatically, is a better predictor of the suppression outcomes of citizens than regime type alone, and that centralized internet exchange points (IXPs) constitute a key technical choke point through which states can control their citizens, regardless of their type of regime. The paper concludes with considerations of the technical and policy compromises between national security goals, platform governance and digital civil liberties.

Keywords: social media surveillance, state control, digital divide, deep packet inspection, internet freedom, content moderation, cross-country analysis, computer science

1. Introduction

Worldwide, social media are a growing part of political communication, civic engagement and information acquisition. This centrality has created social media infrastructure as a highly strategic target for both state actors and groups who want to keep an eye on, influence or silence people



online. In contrast to previous modes of state control, which were mainly by licensing of print and broadcast media, the current social media surveillance is carried out by identifiable computer science tools: network traffic inspection, application programming interface (API) requirements, machine learning content classifiers, and biometric systems of identities. But as well as the legal and political background, a knowledge of the social media surveillance systems architecture is required to understand social media surveillance.

As a system, all such state-level social media surveillance activity eventually involves a degree of control over one of the few layers in the standard networking and application stack: the physical and network layers on which packets are routed and can be inspected or blocked; the transport layer on which encrypted social media sessions are fingerprinted or interrupted; the application layer in which platforms expose APIs and enforce policy; and the data layer in which machine learning models process scaleable volumes of user-generated content. The engineering approach of considering surveillance a stack problem can help in comparing extremely different states on a technical level, and in determining what layer of the stack a specific measure – whether circumvention, platform policy, or legislation – is meant to tackle.

Previous research on Internet censorship has resulted in comprehensive technical quantifications of particular mechanisms—DNS injection or the blocking of specific keywords in URL filtering—whenever it has been attempted, without much connection to other studies or without an agreed-upon framework for comparing the way that the different countries use these mechanisms to create a unified system for monitoring the Internet. A parallel set of cross-national indices like Freedom House's survey Freedom on the Net (FOTN) give comparable outcome scores across a large number of countries, but also report outcome scores as a technical decomposition of the underlying architecture, instead of scores as a composite indicator. This paper seeks to fill that gap by merging the empirical power of the cross-national index approach with the technical taxonomy of systems that can account for the variation in internet freedom outcomes across countries, rather than just measure it.

This paper performs a cross-national technical examination of social media surveillance and the state control. Instead of regarding "surveillance" as one monolithic phenomenon, the paper breaks down surveillance into four technical layers: network infrastructure, platform and identity systems, analytics and artificial intelligence, and enforcement, and a fifth category of analytically distinct, individually targeted "device-level" surveillance, and assesses the implementation of each layer in a set of countries that have seen different outcomes in terms of internet freedom. To serve as an empirical benchmark for comparison, the paper relies on the Freedom on the Net (FOTN) index, a cross-national, annually reported data set that ranks countries based on access challenges,



restrictions on content and violations of user rights, and where available, it pairs FOTN's country narrative with independent peer-reviewed network-measurement research.

This paper has three contributions. First, it suggests a five-category taxonomy of technical dimensions of the state's social media surveillance capacity that can be used in different country contexts and in terms that are not specific to a given country's legal language. Second, it uses this taxonomy to create an eight-country comparative table that connects specific technical mechanisms to their domestic legal basis, as part of the overall FOTN 2022 outcomes. Third, it packages up the case evidence and creates a generalized architecture diagram to assist future more granular network-measurement research.

2. Literature Review

In one of the first systematic cross-national technical surveys of internet filtering, the multi-year OpenNet Initiative testing project reveals that national censorship systems usually rely on multiple blocking methods, e.g., IP blocking, DNS tampering, filtering the URL itself based on keywords, and not any particular method (Deibert et al., 2010). Following this legacy of measurement, Xu, Mao, and Halderman (2011) conducted an empirical study to determine where in the network path the censorship occurs, and found that Chinese censorship occurs in a handful of international gateway routers, not throughout the local network, an early empirical study of the network-chokepoint. Ensafi et al. (2015) took this approach to measurement and applied it in a longitudinal study, which revealed that the behavior of state filtering infrastructures is not fixed, but rather that state filtering infrastructure is iteratively re-engineered to thwart circumvention methods over time. Filastò and Appelbaum (2012) tackled the resulting measurement gap head-on by creating the Open Observatory of Network Interference (OONI), an open-source software platform for distributed, volunteer-operated probing of network interference, which underlies a significant portion of the independent, non-state-reported censorship measurement.

The second strand of research has looked at platform-level content control as a separate problem from network filtering. King, Pan and Roberts (2013) performed an extensive automated content study of Chinese social media posts, and discovered that censorship was not generally exercised on any negative comment about the state, but specifically against posts that could be used for collective action, with direct implications for how the classifiers in the analytics-layer are likely to be tuned in practice. Roberts (2018) later theorized such a pattern as a technique for "porous censorship"—one which works by creating friction, delay and selective flagging, but does not involve outright content removal, and is computationally reliant on classifiers that are able to probabilistically score content. Together, this literature demonstrates that the platform-level



analytics systems are not just a simple keyword filter, but are carefully tuned, at the model-architecture level, to recognize coordination and reach, not content sentiment.

A third thread is about individual surveillance, which is different from the above population-scale mechanisms. The Citizen Lab at the Munk School of Global Affairs, at the University of Toronto, was the first to publicize a forensic analysis of NSO Group's Pegasus spyware, reverse-engineering the attack chain used to spy on a human rights defender's mobile device, and identifying the forensic methodology, device back-up artifacts, network logs, and process traces upon which later Citizen Lab and Amnesty International Security Lab investigations have relied to attribute targeted surveillance incidents across dozens of countries. The paper is limited to the forensic literature in this regard, as is traditional in the field where technical details of exploit levels are not reported in public.

An additional line of research is the study of network-layer shutdown events, rather than of permanent filters. Since 2016, Access Now's #KeepItOn coalition has been keeping a longitudinal, incident-level database of incidents where the internet has been shutdown, categorizing them by trigger (elections, protest, exams), technical method (mobile networks, application layer throttling, national backbone shutdown), and duration, making network-layer shutdown capability a discrete and measurable technical variable, rather than simply a description of the internet being shutdown.

The Freedom on the Net project by Freedom House integrates the results of these four bodies of research into a uniform annual index, which is based on a structured expert evaluation and assigns each of the countries surveyed a score between 0 and 100 for three sub-indices: obstacles to access, limits on content and violations of user rights. The fact that the sub-index structure of FOTN 2022 more closely tracks the network-, platform-, and analytics/enforcement-layer distinctions found in the technical measurement literature reviewed above is important, especially when combined with the cross-country coverage provided by the index in a standardized format, that isn't offered by individual network-measurement studies, which are often single-country or single-mechanism in scope.

Together, this literature points to a gap that is directly addressed by this paper: individual studies have rigorously used specific mechanisms, such as network-path filtering (Xu et al., 2011; Ensafi et al., 2015), platform-level classification behavior (King et al., 2013), targeted device compromise (Marczak & Scott-Railton, 2016), and shutdown events (Access Now); however, there is no widely used framework that combines these mechanism-specific findings to offer a cross-country, comparative architecture.



3. Methodology

The approach of this paper is mixed qualitative-comparative method. The 70 countries covered in the Freedom on the Net 2022 report were narrowed to eight countries that span the spectrum of results for internet freedom: five "Not Free" countries (China, Iran, Russia, Pakistan, and Turkey), two "Partly Free" countries (India, Brazil), and one "Free" country (the United States), serving as a comparative baseline. Countries were also included for evidence of public reports of their use of particular technical surveillance mechanisms, based on country narratives from FOTN, network-measurement studies that have been peer-reviewed, and reports from existing digital-rights research organisations.

For every country, the paper traces the four-layer technical architecture of surveillance—a network infrastructure, platform and identity, analytics and AI, and enforcement—and maps them onto what is publicly known about that country. The mapping is qualitative and descriptive; the paper does not seek to independently verify or reproduce any surveillance technique, nor does it include any detail of operational and/or exploit level technical detail. The numerical scores from the FOTN is used as a 'standard' outcome measure that is independently generated and used to compare against the qualitative technical mapping.

The only sources for qualitative mapping were publicly available documents and peer-reviewed or institutionally published research from trusted network-measurement and digital-rights research bodies, such as the FOTN country narrative reports, published national laws and regulations, and the like. If there was a specific technical claim (e.g., a named DPI system, a specific legal provision permitting network shutdowns) and it was not substantiated by at least one primary legal or institutional source and one independent research or reporting source. The idea behind this approach of source triangulation is to mitigate the danger of overstating the technical surveillance capabilities of a particular country on the basis of a single unverified report.

4. Technical Taxonomy of Social Media Surveillance Mechanisms

Social media surveillance and state control mechanisms can be grouped into four broad technical categories, corresponding to the layer at which they intervene in the flow of information from a user's device to a public audience.

4.1 Network-Layer Controls

Network-layer controls are implemented at internet exchange points (IXPs), at gateway routers and at the internet service provider (ISP) infrastructure. A DPI middlebox filters packets by looking at their headers (and, if unencrypted, payloads) for the presence of traffic for blocked domains,



blocked key words, or signatures of a protocol. When DNS is manipulated such as DNS injection and resolver-level blocking, domain names cannot be resolved to an IP address. Whole address blocks of specific platforms can be blocked using IP-range blocking and BGP-level routing controls. Bandwidth Throttling is a process which slows down access to targeted services, not prevents it, and makes it more difficult to attribute disruption to deliberate state action. At the far end of the spectrum, network-layer "kill switches" enable governments to block mobile or fixed broadband traffic completely in a specific area, as evidenced in numerous FOTN country reports, such as Pakistan's Section 54 of the Pakistan Telecommunications Act, which provides for regionally and event-triggered Internet service shut-downs.

4.2 Platform and Identity-Layer Controls

Platform-layer controls are not direct controls of the network but rather take the form of legal and regulatory requirements placed on social media companies. Examples of such include the "data localization" requirements for local data storage, licensing provisions that restrict legal operation of a platform based on compliance with removal requests from platform operators, and takedown requirements through APIs that enable government-initiated takedowns to be made quickly. Identity-layer controls, such as SIM card registration that is required to be linked to a national identity database and biometric verification requirements, make it less easy to use social media anonymously, strengthening the link between an online account and the verified identity of their real-world users and therefore making it easier to trace posts to their sources.

4.3 Analytics and Artificial Intelligence Layer

The analytics layer is the machine learning and natural language processing systems that are responsible for the scale of the content classification. Text classifiers can identify prohibited terms; image and video hash-matching systems can detect flagged images and videos; sentiment-analysis models can measure the political valence of posts; and social graph analysis techniques can identify influential accounts or coordinated networks, or dissent clusters, that can be detected by interaction patterns, but not contents alone. In certain jurisdictions, platforms must be legally mandated to give state-approved material a higher profile, turning the platforms' own engagement-optimization systems into tools for political messaging, not just reporting.

4.4 Enforcement Layer

The enforcement layer implements the results of the analytics layer in real-time, either automatically or with human review, and takes the form of content takedown, temporary or permanent suspension of accounts, referral to law enforcement and, in the most egregious circumstances, extralegal consequences such as detention. In a number of "Not Free" countries, physical consequences of online speech are also recorded in FOTN country reports, such as arrest,



prosecution under cybercrime laws, and in a few documented instances, enforced disappearance of activists and journalists.

4.5 Targeted Surveillance as a Distinct Modality

Besides the four layers of mass-surveillance, there are some states that have additional targeted technical surveillance specific to persons, usually journalists, opposition political figures, or human rights activists. This is the most basic modality, and is typically associated with commercial mobile spyware that takes advantage of the fact that it doesn't rely on network level chokepoints or platform cooperation for access to messages, location data, and device sensors – no matter what platform or network the person may use in the future. These are tools used in secret, at the target level, and are not documented as mass-surveillance layers, except by means of forensic analysis of the compromised devices. In this paper, the concept of targeted surveillance is analytically separate from but possibly allied to the four-layer model of mass-surveillance architecture, which is a state with limited mass network-control infrastructure but might still have a very capable targeted-surveillance program.

4.6 Infrastructure Capacity as a Determinant of Surveillance Reach

There are important structural implications in the technical taxonomy developed above: each of the four layers of mass-surveillance needs a basic level of infrastructural and institutional capacity that is unevenly distributed between countries. There are multiple, fragmented gateway infrastructure providers in countries with fragmented gateway infrastructure, while in countries with a smaller number of state licensed IXPs, it is easier to have centralized control of international gateway infrastructure for the national scale deployment. AI-content categorization demands ongoing investments in machine learning engineering capabilities, annotated training data, and GPU-scale computing hardware, which are all centralized in a few highly resourced states and platform companies. Biometric identity-linkage systems depend on a national digital identity system, which in turn relies on previous investments in national identity databases, enrollment hardware and the infrastructure to reach enrollment points. Thus, the technical range of a state's surveillance system, i.e. which of the layers can be implemented in practice, depends on the same variables in terms of infrastructural capacity (bandwidth, computing capacity, centralized network topology) that affect its overall level of ICT development, regardless of its legal and political structure.

4.7 Encrypted Traffic Analysis and Circumvention-Resistant Filtering

The recent adoption of Transport Layer Security (TLS) for web and application traffic has driven the change in the network layer filtering industry from just dealing with payload inspection. TLS encrypts the contents of HTTP requests, which means that a growing number of DPI systems are



based on information that is visible in the TLS handshake, for instance the Server Name Indication (SNI) field sent in cleartext at the beginning of the TLS handshake, which is transmitted even if the following traffic is encrypted and indicates a client's intention to connect to a specific domain. The literature review shows national filtering systems documented in the literature that can block connections by matching against this SNI field without the need to decrypt traffic and run at line rate on commodity middlebox hardware. To remove this metadata leak, protocol-standards-level extensions to the handshake, called Encrypted Client Hello (ECH), have been developed, with some of the network-layer control regimes reported to block or downgrade connections that attempt to negotiate extensions such as ECH, reflecting an emerging tension between protocol designers and state filtering infrastructure.

A related technique is called traffic fingerprinting which involves the use of statistical machine learning models, such as a support vector machine or, in more recent work, a convolutional and recurrent neural network (CNN), to identify traffic flows from encrypted traffic without the need to examine the payload contents. It enables a filtering system to have a good chance of identifying traffic as belonging to a particular application—such as a circumvention tool like Tor—or commercial VPN protocols, even when the traffic itself is encrypted and no information in plain text is carried as metadata. Website fingerprinting research has shown that these classifiers can attain high accuracy to classify application traffic in an extremely constrained laboratory setting, just from traffic traces from encrypted web traffic, which implies that application traffic can be made statistically indistinguishable from typical encrypted web traffic with the development of pluggable transport protocols, such as packets-padding and traffic-shaping obfuscation layers. This is an adversarial pattern, of classifiers vs obfuscation, and serves as an example of how this technical battle between state surveillance infrastructure and circumvention engineering can happen at pretty much every layer of the taxonomy.

4.8 Machine Learning Architectures Used in Content Classification

In deployments today, the analytics-layer classifiers are mostly based on one of three machine learning architecture families. Platform-level content control has evolved from static lists of keywords to transformer-based language models, which have been trained on political-speech corpora labeled by the local community, as the latter perform better in the face of unseen content when using a different phrasing; one can also use the language model for text classification and sentiment scoring. Perceptual hashing systems based on convolutional neural networks, like the Microsoft PhotoDNA, which is well known in the public domain, create short, numerical hash keys of image and video frames that are resistant to changes and alterations, and thus enable identification of previously flagged visual content on re-upload without the need to store or re-inspect the original file. The social graph analysis level employs graph neural network and community-detection algorithms to detect clusters of coordinated accounts or influential nodes



from interaction data (follows, retweets, co-posting timing) without relying on the content which the accounts create, as this is the computational basis for the detection of coordinated, and not just individually objectionable, activities.

5. Cross-Country Comparative Analysis

Figure 1 presents Freedom on the Net 2022 scores for the twelve countries used as reference points in this analysis, illustrating the wide spread between "Not Free" states scoring below 40 and "Free" states scoring above 70. China's score of 10 reflects, according to the FOTN 2022 country report, the combination of a centralized DPI-based national firewall, mandatory real-name registration, and legal requirements that platforms tune recommendation algorithms to promote state ideology. At the opposite end, the United States and United Kingdom score in the high 70s, reflecting comparatively limited platform-level content mandates and surveillance activity constrained by judicial warrant requirements, notwithstanding documented concerns about metadata-based lawful-intercept programs.

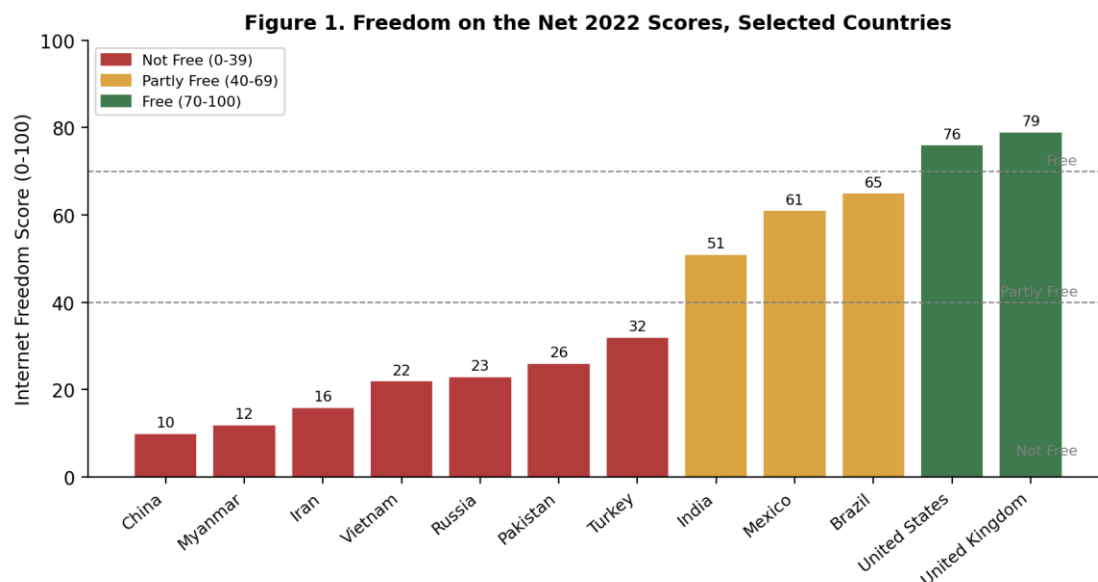


Figure 1. Freedom on the Net 2022 scores for twelve reference countries, colored by status band. Source: Freedom House, Freedom on the Net 2022.

Figure 2 situates these country-level scores within the global trend documented in the FOTN 2022 report: of 70 countries surveyed, internet-freedom conditions declined in 28 and improved in 26 during the coverage period, continuing a trend of overall global decline. This distribution indicates



that deteriorating conditions were not confined to a small set of outlier states but represented a broadly shared trajectory across the sample.

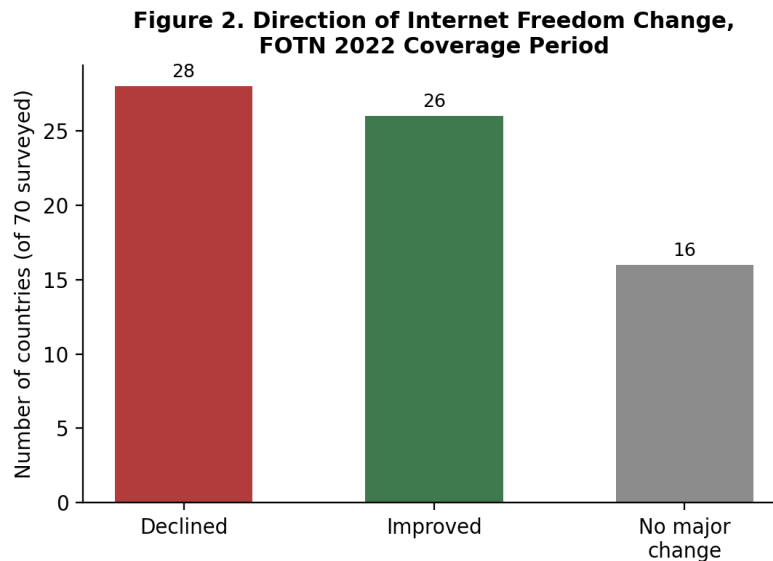


Figure 2. Direction of internet freedom change across 70 countries surveyed in Freedom on the Net 2022. Source: Freedom House.

Table 1 maps each of the eight case countries onto the technical taxonomy, alongside the primary domestic legal or institutional basis for the documented surveillance techniques. The table illustrates that lower FOTN scores are consistently associated with a broader combination of techniques spanning all four layers of the taxonomy, whereas higher-scoring countries tend to rely on a narrower set of techniques concentrated at the enforcement layer, typically constrained by judicial process.

Country	FOTN Status 2022	Primary Surveillance Techniques	Legal / Institutional Basis	FOTN Score
China	Not Free	National DPI firewall ("Great Firewall"), mandatory real-name registration, AI content classifiers tied to ideological-promotion mandates	Cybersecurity Law; Data Security Law	10
Iran	Not Free	National Information Network ("halal internet"), platform blocking, state-run app substitutes, periodic shutdowns	Computer Crimes Law	16
Russia	Not Free	SORM lawful-intercept system,	Sovereign Internet	23



		mandatory DPI equipment (TSPU), sovereign RuNet routing, platform bans	Law; Yarovaya Law	
Pakistan	Not Free	PTA-mandated blocking/shutdowns, content-removal rules, SIM-to-CNIC biometric registration	PECA 2016; Removal and Blocking of Unlawful Content Rules 2021	26
Turkey	Not Free	Platform representative-office mandates, bandwidth throttling, court-ordered takedowns	Law No. 5651; Social Media Law 2020	32
India	Partly Free	IT Rules intermediary compliance, SIM-linked biometric ID, periodic regional shutdowns	IT Act 2000; IT Rules 2021	51
Brazil	Partly Free	Judicial content-removal orders, data-retention mandates, limited platform-level surveillance	Marco Civil da Internet; LGPD	65
United States	Free	Lawful-intercept metadata programs, platform self-regulation, targeted judicial warrants	CALEA; FISA	76

Table 1. Cross-country comparison of documented social media surveillance techniques and their legal basis, Freedom on the Net 2022 coverage period.

6. Case Studies

6.1 China: Centralized Network Control and Algorithmic Mandates

China's surveillance system is the integration of the 4 layers in a centralized system. A national DPI infrastructure, typically known as the "Great Firewall", does inspection and filtering at international gateways at the network layer. Real-name registration is required for social media accounts at the platform and identity layer level, which means linking online activity to verified national identity records. At the analytics level, FOTN 2022 registers new laws that require social platforms to adapt their algorithms to boost content that "conforms to the line of the ruling party," as opposed to just their moderation system. At the enforcement level, penalties for unauthorized online speech are criminalization based on cybersecurity and public-order laws.

6.2 Russia: Sovereign Internet Infrastructure

The Russian Internet Law of the state requires that domestic Internet companies equip their network with government-provided DPI equipment (Technical Means for Countering Threats, or TSPU), which can filter and, when necessary, isolate Russian networks from the global Internet.



Together with SORM, a longstanding system of lawful interceptions where ISPs are directly vulnerable to government access to communications data, this forms a network-layer surveillance system that won't depend on individual platform cooperation. During the reporting period covered by FOTN 2022, Russia stepped up its blocking of platforms, specifically platform-level blocking of major social media platforms, which shows how network-layer blocking and platform-layer blocking can be implemented simultaneously, when the political climate is deemed to be acute.

6.3 Pakistan: Regulatory and Shutdown-Based Control

The surveillance and control framework in Pakistan is more about maintaining regulations and shutting down networks as compared to a large centralized DPI system as in China or Russia. The Pakistan Telecommunication Authority (PTA) has the authority to suspend networks under Section 54 of the Pakistan Telecommunications (Re-organization) Act, which is a power that has been being exercised routinely in the country even outside of the originally intended emergency context as documented in the country report of the FOTN 2022. The Removal and Blocking of Unlawful Content (Procedure, Oversight, and Safeguards) Rules were introduced in the coverage period of 2022. In the identity layer compulsory registration of the SIM gives away the anonymity of access to mobile based social media, this is the most common use of the internet for most Pakistani users.

6.4 United States: Judicially Constrained Surveillance

The U.S. case represents a more limited form of surveillance, which is focused at the enforcement end and limited by statutes, including the Foreign Intelligence Surveillance Act (FISA) and the Communications Assistance for Law Enforcement Act (CALEA). Platform-level blocking of domestic platforms is very limited, and there is little government intervention in platform-level content moderation, with most happening under private terms of service, although some is currently being discussed due to concerns about the extent of private self-regulation in the context of election-related misinformation (FOTN 2022). As such, in the present analysis this case serves more as a comparative baseline and as a representation of the same technical capabilities as discussed elsewhere in this document, namely network-level traffic inspection, but in a more restricted configuration, under a different legal and institutional framework.

7. Generalized Technical Architecture of State Surveillance

Synthesizing the case evidence, Figure 3 presents a generalized four-layer technical architecture describing how state social media surveillance systems are typically composed, regardless of the specific country context. Traffic and identity data collected at the network and platform layers feed into the analytics layer, whose outputs in turn determine which content, accounts, or individuals are referred to the enforcement layer. This generalized architecture is not intended to describe a



specific deployed system in any one country, but to provide a common analytical frame for comparing partial implementations across different political and institutional contexts.

Figure 3. Generalized Technical Architecture of a State Social Media Surveillance Stack

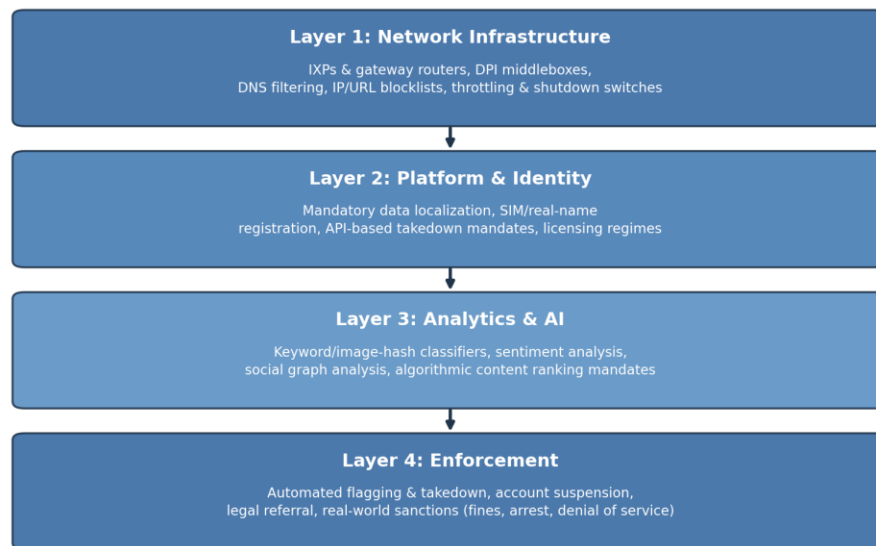


Figure 3. Generalized technical architecture of a state social media surveillance stack, synthesized from documented country practices.

A notable pattern across the case studies is that centralized internet exchange points function as a key technical chokepoint regardless of regime type: because the large majority of international social media traffic in each case country passes through a small number of state-licensed or state-adjacent gateway providers, network-layer control at this chokepoint offers disproportionate leverage relative to its infrastructural cost. This suggests that the technical architecture of a country's internet backbone, specifically, the degree of centralization at its international gateways, is itself an important explanatory variable for surveillance capacity, independent of the country's stated legal framework.

8. Discussion and Limitations

There are a number of limitations to this analysis. The first is that FOTN scores are not technically measured network traffic; instead, they are generated using expert coding, which can also have



methodological limitations such as different weights assigned to similar methods across countries. Second, publicly reported surveillance methods are only likely to underestimate the level of capacity state monitoring powers have, as covert powers are inherently hard to prove through open source reporting. Third, the four-layer taxonomy makes it easy to describe and understand overlapping and incremental systems in practice; a single piece of technology (e.g., a required government-sanctioned DPI appliance) can perform both network-layer filtering and analytics-layer data-collection functions. Fourth, the country selection in this paper covers the full range of FOTN status, but is not a random or representative sample, and should not be interpreted as causal relationships between certain technologies and internet-freedom outcomes.

However, given these constraints, the technical mapping provided here offers two policy-relevant lessons. First, network-layer chokepoint control, not just platform-level content moderation, seems to be the more meaningful technical lever that states have to gain the comprehensive surveillance and control ability that they desire, with ramifications for the prioritization of digital-rights interventions. Second, the growing demand from several jurisdictions to open up administrative APIs or adjust ranking algorithms to government's specifications is a form of state involvement in a system that in past periods of Internet governance studies were assumed to be privately governed, beyond direct state technical reach.

The results also have implications for the design of resilience mechanisms on the network and application levels. Circumvention technologies work by trying to shift traffic into a different part of the country's technical architecture; how well this works depends on which layer of the surveillance structure the country uses the most. Circumvention tools that focus on network-layer DPI filtering, for instance, provide some protection for platform-layer identity linkage or content classification of analytics-layer content, in that a properly authenticated, real-name-registered account is still accessible to platform-level enforcement tools from both network-layer and platform-layer paths. This point indicates that digital rights interventions may be best targeted at the layer—or layers—of a state's surveillance apparatus it depends on, as opposed to being applied as a standard technical solution everywhere.

Last, but not least, is a normative consideration that should be explicitly formulated and discussed beyond the technicalities. The individual technical mechanisms, such as lawful-intercept systems, content-moderation classifiers and identity verification, also form the basis for legitimate public-safety, counter-fraud, and platform-integrity capabilities in states with robust judicial oversight. This distinction between country cases that this paper has made is not simply whether or not a given technology is being used but whether its use is constrained by independent judicial process, whether it is subject to public accountability mechanisms, and whether it is proportionate to a specific lawful and clearly stated purpose in the enforcement or security sphere as reflected in the FOTN sub-score on "violations of user rights" used in this analysis.



9. Conclusion

This paper has provided a cross-country technical analysis of social media surveillance and state control, divided into four layers to describe network infrastructure, platform and identity systems, analytics and artificial intelligence, and enforcement, with an analytically separate fifth layer being individual targeted surveillance. The extent of technical layers subject to state control, not just any single technology, appears to be strongly correlated with the internet-freedom outcomes in eight countries studied, though centralized network chokepoints seem to be a common enabler of state control across very different political environments. The analysis also shows that the process of digital-infrastructure development (as a whole, not just the capacity to build surveillance systems) is not separate from the process of closing digital divides in service delivery: The same kind of centralized ICT infrastructure investments that can be used for closing digital divides in accessing services can also be used for monitoring populations on a vast scale, if there are no independent constraints on how these systems are used by judiciary and other institutions.

References:

- Freedom House. (2022). *Freedom on the Net 2022: Countering an Authoritarian Overhaul of the Internet*. Washington, DC: Freedom House.
- Freedom House. (2022). *China: Freedom on the Net 2022 Country Report*. Washington, DC: Freedom House.
- Freedom House. (2022). *Pakistan: Freedom on the Net 2022 Country Report*. Washington, DC: Freedom House.
- Freedom House. (2022). *United States: Freedom on the Net 2022 Country Report*. Washington, DC: Freedom House.
- Freedom House. *Freedom on the Net Research Methodology*. Washington, DC: Freedom House.
- Deibert, R., Palfrey, J., Rohozinski, R., & Zittrain, J. (Eds.). (2010). *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. MIT Press.
- Xu, X., Mao, Z. M., & Halderman, J. A. (2011). *Internet Censorship in China: Where Does the Filtering Occur?* In *Proceedings of the Passive and Active Measurement Conference (PAM 2011)*. Springer.
- Ensafi, R., Winter, P., Mueen, A., & Crandall, J. R. (2015). *Analyzing the Great Firewall of China Over Space and Time*. *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2015(1).
- Filastò, A., & Appelbaum, J. (2012). *OONI: Open Observatory of Network Interference*. In *Proceedings of the 2nd USENIX Workshop on Free and Open Communications on the Internet (FOCI '12)*.
- King, G., Pan, J., & Roberts, M. E. (2013). *How Censorship in China Allows Government Criticism but Silences Collective Expression*. *American Political Science Review*, 107(2), 326-343.
- Roberts, M. E. (2018). *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton University Press.
- Marczak, B., & Scott-Railton, J. (2016). *The Million Dollar Dissident: NSO Group's iPhone Zero-Days Used Against a UAE Human Rights Defender*. Citizen Lab, Munk School of Global Affairs, University of Toronto.



Access Now. #KeepItOn: Internet Shutdowns Tracker and Annual Report.

Digital Rights Monitor. (2022). Pakistan Sees No Significant Progress on Internet Freedom, Report Highlights.

Government of Pakistan. Pakistan Telecommunications (Re-organization) Act, 1996, Section 54.

Government of Pakistan. Removal and Blocking of Unlawful Content (Procedure, Oversight and Safeguards) Rules, 2021.

Microsoft. PhotoDNA: Perceptual Image Hashing for Detection of Known Illegal Imagery. Technical Overview.

Rescorla, E., Oku, K., Sullivan, N., & Wood, C. A. (2023). TLS Encrypted Client Hello. Internet-Draft, Internet Engineering Task Force (IETF).