To appear in July 2023 issue of ACM Computer Communication Review

Report of 2021 DINRG Workshop on Centralization in the Internet

Christian Huitema
Private Octopus
USA

Geoff Huston
APNIC
Australia

Dirk Kutscher
HKUST(GZ)
China

Lixia Zhang
UCLA
USA

This article is an editorial note submitted to CCR. It has NOT been peer reviewed.
The authors take full responsibility for this article's technical content. Comments can be posted through CCR Online.

ABSTRACT

The Internet Research Task Force (IRTF) Research Group on Decentralization of the Internet (DINRG) hosted a workshop on *Centralization in the Internet* on June 3, 2021. The workshop focused on painting a broad-brush landscape of the Internet centralization problem space: its starting point, its driving force, together with an articulation on what can and should be done.

CCS CONCEPTS

• **Networks** → **Network protocol design**; **Network security**; **Denial-of-service attacks**.

1 INTRODUCTION

There is a consensus within the networking community that the Internet consolidation and centralization trend has progressed rapidly over recent years, as measured by the structural changes to the data delivery infrastructure, the control power over system platforms, application development and deployment, and even in the standard development efforts. This trend has brought impactful technical, societal, and economical consequences.

When the Internet was first conceived as a decentralized system 40+ years back, few people, if any, could have foreseen how it looks today. How has the Internet evolved from there to here? What have been the driving forces for the observed consolidation? From a retrospective view, was there anything that might have been done differently to influence the course the Internet has taken? And most importantly, what should and can be done now to mitigate the trend of centralization? Although there are significant interests in these topics, there has not been much structured discussion on how to answer these important questions.

The IRTF Research Group on Decentralization of the Internet (DINRG) organized a workshop "Centralization in the Internet" on June 3, 2021, with the objective of *starting* an organized open discussion on the above questions [1]. Although there seems to be an urgent need for effective countermeasures to the centralization problem, this workshop took a step back: before jumping into solution development to steer the Internet away from centralization, we wanted to discuss *how* the Internet has evolved and changed, and *what* have been the driving forces and enablers for those changes. The organizers and part of the community believe that a sound and evidence-based understanding is the key towards devising effective remedy and action plans. In particular, we would like to deepen our understanding of the relationship between the architectural properties and economic developments.

This workshop consisted of two panels, each panel started with an opening presentation, followed by panel discussions, then open-floor discussions. There was also an all-hand discussion at the end. Three hours of the workshop presentations and discussions showed that this Internet centralization problem space is highly complex and filled with intrinsic interplays between technical and economic factors.

This report aims to summarize the workshop outcome with a broad-brush picture of the problem space. We hope that this big picture view could help the research group, as well as the broader IETF community, to reach a clearer and shared high-level understanding of the problem, and from there to identify what actions are needed, which of them require technical solutions, and which of them are regulatory issues which require technical community to provide inputs to regulatory sectors to develop effective regulation policies.

This report roughly follows the structures of the panels. §2 summarizes Panel 1's opening presentation by Geoff Huston, which shows convincingly that, from the industrialization history to today's Internet, economies and opportunities of scale drive industry players towards consolidation of resources and control. §3 summarizes Panel 2's opening presentation by Christian Huitema, which elaborated on several specific factors that played a role in the Internet's evolution from its initial decentralized rollout to where we are today. Because the discussions at Panel-1, Panel-2, and the all-hand discussion at the end of the workshop covered broad and overlapping topics, we summarize all the discussions in §4. In addition to provide a faithful summary of the workshop and to reflect diverse views of the participants, we also inject observations and commentaries at various places, with a hope to offer additional food-for-thought and to suggest potential topics for future studies.

The summary section (§5) captures the workshop's major observations from the Internet centralization problem space, and articulates the lessons to be learned and new insights to be derived. We note that Internet consolidation has been observed across all aspects of the Internet, ranging from network access to infrastructure and to application deployment. The concern regarding centralization is not solely about the size of today's cyber giants, but rather, *the control power and the influence* they possess over users, and over society as a whole. The observed centralization is the result of unregulated economy of scale; technology alone is not responsible for the observed consolidation and centralization, nor can it offer effective mitigation solely.

We further note that society thrives on the balance between economy, regulation, and technology. Today, we see an imbalance which is tilted to economy, with the regulations facing challenges of figuring out exactly what should be regulated and how, and the technologies falling behind security threats; the lack of effectively solutions to mitigate security threats and enable distributed applications is one of the major contributing factors to unwanted centralization of the Internet infrastructure.

Finally, we note that this report is 2-year overdue, and new sea changes have occurred over this time period. In particular, OpenAI with associated GPT transformation are benefiting from multi-billion dollar capital investments and brining disruptive effects that are yet to be fully comprehended. Nevertheless, we believe that the lessons and insights reported herein remain unchanged, and could contribute to future open discussions to move the community to a shared understanding on the problem space, from there to derive next steps towards the Internet decentralization.

2 PANEL 1 PRESENTATION: IS BIG NECESSARILY BAD?

Geoff Huston offered the opening presentation for Panel 1 on "*Is 'Big' Necessarily 'Bad'?*" (see [2] for the presentation slides). He drew lessons from history, using the US economy as an example to articulate the root cause of centralization. About one and half centuries ago, the US went through a phase of industry centralization which was enabled by a liberal labor market, large capital funding (from Europe), and the opening of the railroads for transportation which transformed many smaller regional markets into a single large national market. The impact of the resulting industrial superpowers went well beyond national boundaries over the ensuing decades. A few large-scale US enterprises dominated the emerging global market, and the US domestic economy was a major beneficiary of this position through much of the twentieth century. Their rapid expansion also overwhelmed government regulatory measures at the time, allowing these emerging enterprises to script their own preferred regulatory ruleset. When the industrial incumbents get to define the terms of trade, the inevitable outcome is the suppression of competition, and the transformation of dominant positions by such actors into entrenched monopolies.

The well-known 1890 Sherman Anti-Trust Act authorized US federal government to prohibit companies from colluding or merging to form an effective monopoly, to help workers and smaller enterprises by encouraging competition. The law was applied in 1910 to Standard Oil, American Tobacco, and General Electric to curb their concentrations of power. However, breaking up these monopolies led to some unintended consequences, including national economic panic and a depression in the following year, and the creation of vigorous political lobbying by the large industrial enterprises to bring their interests to the political process and to bind the interests of politicians to their interests.

Huston went on to explain that today's Internet industries seem following the footstep of the past giants, with massive concentration in their chosen areas of activity and massive lobbying. One big difference is that, while the past giants exploited the labor force to accumulate profits, today's Internet giants monetize the accumulated large pools of personal data via advertising markets. With today's Internet practice, individual users simply have no means to profit from their own personal data, even if they would like to. The data is intrinsically valuable in terms of its aggregate volume in conjunction with its individual specificity, which implies that only large enterprises can amass such a critical mass of profile data. This has created a positive feedback loop, where the accumulated value of personal data can underwrite the investment in services that are offered to users at heavily discounted prices, or more commonly as free service, which in turn attract more users and enable further data harvesting. These popular and free services lead to further service concentration. The wealth of these enterprises lies not in the value of physical goods, nor in the value of the services they provide, but in the volume and accuracy of the user preference data that they have amassed. A term "surveillance capitalism" has been coined to describe this new enterprise model [3].

To date, the Internet giants have benefited from the lack of appropriate resolution of data security and privacy concerns. The absence of regulatory imposts regarding data security and due attention to user privacy concerns resulted in the absence of efforts to safeguard user interests on the part of these enterprises. The EU's GDPR effort [4] represents a welcome change in this regard, however the financial fines used in this measure may be too low to curb the behavior of large enterprises. At the same time, such punitive regulatory measures could bring potential unintended consequences of disincentivizing smaller new entrants in the market and further entrenching the position of the largest incumbents.

Huston also alerted people of an even bigger danger, the concentration in the web search market, where the dominant player occupies 92% of world-wide search market at this time.[1] Search has replaced all forms of reference libraries. Search engine has become the ultimate decider of public argument, the ultimate tool for researchers, the ultimate source of all information. Whenever one has a question, no matter what kind, "*Google answers all questions. Google tells one what to think*". Dominant search engines are both incredibly empowering, allowing the world's knowledge at one's fingertips, and incredibly threatening as they can shape society's views in all subject areas, from the fashions of the day to world politics.

Observing the creation processes of industrial giants from history to present, one may conclude that the appearance of dominant industrial players is a natural outcome of the compounded gains from economy of scale. In addition, big players grow way faster than public's comprehension of their consequences, let alone the regulation development. When a market player grows big enough, it becomes possible for it to make its own rules.

The above pointed out the negative consequences of being big. However, Huston also reminded us that being big is not *all* bad. Big players can bring national economic gains. They also possess big capital to invest into future technology advances, thus prohibiting companies of big sizes may destroy potential opportunities. In short, being big can have positive effects, and breaking big companies may lead to economic pains.

On the other hand, enterprises seek to maximize their own profits, and shareholders' interests do not match public interests in general. Therefore, the public sector needs to regulate big players to prevent their excessive aggregation of power from interfering with competition and consumers interests. It remains a big challenge to (1) determine *what* and *how* to regulate to allow society to

[1] https://gs.statcounter.com/search-engine-market-share

both minimize the economic impact of the regulations and maintain the benefits from economy of scale, and (2) to gain a clear understanding of *what results* one may expect from the defined regulations.

3 PANEL 2 PRESENTATION: HOW DID WE GET HERE?

Christian Huitema presented the opening talk in Panel 2 on "*How Does Centralization Start?*" (See [5] for thte presentation slides). He pointed out that the Internet centralization trend started many years back with companies investing efforts into the new area of Internet applications to meet users' needs. As the Internet growth exploded during the 90's, information search became an immediate need. In response, early attempts include Archie, Altavista, and *Yahoo!* to meet users' needs. But a relative latecomer, Google, quickly developed new search technologies and advanced past the competition. There were also plenty of early diverse efforts on user identities, contacts, and presence, they were overtaken by centralized services such as AOL, MySpace, and eventually Facebook won the race as the dominant "social network" service provider.

While search and social networking applications started with centralized service providers, email, which was the first widely used application on the Internet, started with a predominantly distributed deployment based on federated email servers. Although a few Internet portal service providers, such as AOL and Yahoo!, offered email services to individual users, most institutions and companies deployed their own email servers. However, over time, Google, as an emerging dominating IT company, made significant investments in developing and delivering a popular email service. Gmail leveraged its scale to effectively combat spam, and ultimately gained a significant share of the email service market with their Gmail product in many regions.

Web hosting and DNS services are two other similar cases, with many institutions and companies running their own web servers and DNS servers at the beginning. Again, as time went on, increasing network security threats, such as viruses and DDoS attacks, together with shortage of IT manpower, resulted in web services being consolidated by the big cloud providers, AWS, Azure and Google cloud, and a large fraction of DNS authoritative services is now provided by Cloudflare, AWS, and GoDaddy. Big web hosting companies can leverage competitive advantage from economies of scale and can invest in additional capacity to mitigate DDoS through building castles with strong walls (well protected data centers). Enterprises today feel very exposed if they set up their own online servers as a critical part of their IT service offerings, as they could be brought to a halt by a DDOS attacks, not to mention possible compromises of the server software and user data.

Once those early commercial application providers started their business, multiple factors have driven them to move their offerings up the protocol stack and centralize further. First, it is costly and time consuming to deal with multiple operating systems, multiple versions to coordinate development efforts. The problem becomes much easier if one can i) develop applications running on centralized servers, and ii) control the client platforms, which enables one to simply ship the application code whenever a new feature is added.

Second, the emergence of big data and machine learning has facilitated and accelerated centralization. Some service providers (SPs) have started with free services to attract users. To serve users well requires knowing them well. Gaining more users leads to bigger data collections, which enables development of better services, hence more customers, creating a positive feedback circle: more users $\Longrightarrow$ more data for ML $\Longrightarrow$ improved services $\Longrightarrow$ more users and higher revenues. One example that shows the importance of big data is Bing's initial lack of success, despite Microsoft's big investments into its development and deployment. Possessing user data, that a newcomer did not have, enabled Google to tailor search results for individual users in a way that Bing could not match.

On the flip side, many of these large-scale free services are financed by (targeted) advertising, which in turn requires that the advertisement platform amass a collection of user profiles to gain the ability for targeted advertisement to maximize the effectiveness. The proliferation of such free services leads to greater reliance on a surveillance economy to sustain this business model. The more these SPs know about individual users, the better services they can provide, but also the more control they have over the users through selective information provision, blurring the line between services and influence. Recent years have seen plenty of evidence of such influence as documented by Zuboff in "The Age of Surveillance Capitalism" [3].

Huitema observed that the playground of centralized services has changed over time. For example, Microsoft Windows used to dominate the desktop market and through that a lot of application markets, but it is no longer the case. One might explain this by a few relevant factors. One is that the company missed the smart phone market, which broke the lock Microsoft had on applications; another is the lawsuit by the US government, which forced Microsoft to abide to a "consent decree". It had to publish APIs to promote interoperability with other systems, say enabling people to develop their own version of Powerpoint – a decision that seems to have weakened Windows' dominance.[2] However, one can also see the consolidation of the overall OS market by moving to open-source OSs such as Linux in some areas, except the OSs for phones which settled with Android[3] and iOS.

Huitema concluded his presentation by pointing out that decentralized competitors can face uphill battles against centralized providers. Although surveillance capitalism exploit user privacy to fund free services and attain monopoly positions to maximize profits, the solution space seems largely to be a political one. Furthermore, decentralized competition requires standards, and standard development is costly in both efforts and time. In contrast, it is far easier, simpler, and faster for monopoly service providers to develop new applications and to add new features.

4 WORKSHOP DISCUSSIONS

We sort the workshop discussions into three different aspects of centralization:

[2] This example shows that publishing APIs could be a worthwhile issue to consider in future decentralization efforts. The same suggest is made in [6].

[3] Android is based on Linux and thus uses GPL for the OS kernel. The user space Android software is licensed under Apache GPL Version 2.0 (https://source.android.com/docs/setup/about/licenses). Overall, the Android system itself has largely consolidated upon a collection of open-source software components.

- What are the different relevant aspects of centralization (*Categorizing Centralization*)?
- What factors cause or contribute to centralization (*Factors of Centralization*)?
- What are the future perspectives (*Looking Ahead*)?

We also added editorial comments in italics, marked with the word "*Comment:*".

4.1 Categorizing Centralization

Several workshop attendees brought up the notion that centralization can be sorted into multiple categories:

- Operational centralization, as we have described.
- Development consolidation, as measured by the people and organizations that are developing network protocols. Today there seems to be a very small number of organizations, concentrated in a few countries that are developing Internet protocols.
- Centralization and consolidation at different protocol layers. For example, consolidation of transport protocols (e.g., QUIC), consolidation of DNS services, and of course consolidation of applications and services such as e-mail.

On the other hand, some participants pointed out that the passing years have seen some fundamental changes in the value chain. Networking started from dominance of carriage, then moved to dominance of platforms, and today's dominance is application services. As technologies advance over time, lower layer services became commodity services, and the locus of value and money are shifted up the protocol stack, where one can exploit centralization with minimized cost. Today, applications themselves take over the control of everything, creating their own ecosystem. Thus, centralized control of DNS operations and QUIC development could be viewed as part of that ecosystem. Therefore, the observed consolidation symptoms in different categories may in fact all share the same root cause.

4.2 Factors of Centralization

Four different factors of centralization have been discussed:

- Centralization driven by economies of scale;
- Applications got centralized; security threats further intensified centralization;
- Internet communication model: anyone can send to anyone else; and
- the failing of network security.

It was noted that in some areas such as network security, platform-based services and their centralization addressed (or exploited) certain shortcomings and operational complexities – on the other hand the concentration into platform monocultures also created new threats because single attack vectors could affect large user groups.

4.2.1 Centralization Driven by Economies of Scale.

The workshop participants largely agreed that economies of scale have played a center role in the Internet consolidation. In the absence of regulatory restrictions, markets naturally consolidate when economies of scale come into play. More specific driving factors for this consolidation include:

- The economies of scale enable one to generate the same service outcome with far lower production costs, requiring fewer resources for each instance of the service transaction.
- A large user pool produces big data, which helps improving service customization for each user, letting bigger companies gain an edge over smaller competitors.
- Centralized application developments reduce the number of platforms, hence substantially reduce the cost in development and maintenance, and circumvent interoperability issues. In addition, consolidated development and operational efforts also help mitigate technical expertise shortages.
- Monopoly players have the power to control both the terms of service and the price of their offerings in the market. This dominance can lead to prolonged stagnation and reduced efficiency within the market, hindering further innovation.

4.2.2 Applications Got Centralized; Security Threats Further Intensified Centralization.

Both panel speakers pointed out that the playground of networking has changed in fundamental ways, driven by multiple factors. As we discussed already, economy of scale is the most important driver. At the same time, we should not overlook network security, or lack of it, as another big factor that has contributed to application and service centralizations.

The TCP/IP protocol stack did not come with security built-in, and naïve IP devices with implementation and configuration weaknesses can trivially be compromised, creating a fertile ground for malign exploitation. Universal IP connectivity has been massively abused by DDoS attackers. Thus, big players have built fortress, the cloud services, with strong walls to fence off DDoS attacks. All end users are now required to connect to clouds via secure connections, resulting in today's networking picture, looking from 3,000 feet, remotely mimicking the one from the 70's, where all user terminals connected to main frame computers via dial-up lines.

By and large, today's network applications run in clouds; few, if any, run over direct user-to-user or device-to-device communications, with Apple's Airdrop as a noticeable exception.

> *Comment: Apple's Airdrop runs over secure connections between iDevices because Apple installs user identities and corresponding crypto credentials on those devices. In today's Internet, generally speaking, only servers have DNS-named identities and Certificate Authorities (CAs) assigned crypto credentials, but not end users and user devices. Users have application provider assigned identifiers, such as Gmail addresses or Facebook IDs. Thus, they can only be authenticated by those application providers. Consequently, the only means of secure user-to-user communication is through cloud-based authentication services.*

4.2.3 Regarding "Allowing Anyone to Send to Anyone Else".

One debated question at the workshop discussions is whether IP's original model of "enabling any host to send packets to any other host" is one of the major causes of today's security threats. Indeed, IP enabled packet exchanges between any host pairs, similarly TCP/UDP lets any process connect to any other process. In its early days, this universal reachability model enabled the Internet to grow rapidly and spread to the entire world. Today, unfortunately, miscreants use this same feature to launch DDoS attacks at global scale, and email spams and phishing attacks fill the majority of today's total e-mail traffic.[4]

The workshop participants pointed out that source IP addresses can be easily spoofed today, enabling reflection DDoS attacks and making it difficult to traceback attackers. Similarly, there is also no wide, effective solution deployment to shut down spammers. Not being able to clearly identify attack sources, i.e., who are sending those unwanted traffic [7], is a problem that we have not worked out effective fixes.

> *Comment: We do not believe that the value of universal reachability has changed over time. What has changed is TCP/IP's operational environment, from being a playground for research community to political and economic battlefields of worldwide scale. It is not the universal reachability model, but the* lack of security in networking*, that pushed defenses to higher layers and mandates cloud-based solutions, per-service fortresses, and consequent centralization that we see today.*

4.2.4 Network Security Is Failing.

The above discussion reflects the ineffectiveness of existing security solutions as of today, which is also one of the major challenges raised by the workshop submissions.

In general, security threats are a broad topic that exhibit at different layers in the Internet. Among a wide range of different types of attacks, two most noticeable ones are volumetric DDoS attacks at network layer and security breaches, including data theft, at higher layers. Devices deployed long time ago, but still online today, exhibit blatant vulnerabilities; and lack of security in vanilla TCP/IP implementations, compounded with misconfiguration, on inexpensive IoT devices resulted in massive device compromises and global-scale DDoS attacks [8]. The scale of attacks overwhelms the protection capabilities of individual organizations, driving them to outsourcing. Large IT companies, such as Cloudflare, Microsoft, Google etc. to name a few, can afford significant capital investments for building fortresses to protect their assets and take in their customers.

Today, even with wide deployment of Transport layer security (SSL, TLS), insecure software packages and insecure configurations in many deployed systems can also lead to large scale system compromises [9]. Transport layer security and generally better security practices do not prevent all attacks, and we are observing permanent arms races between software providers and IT management on one side and resourceful attackers on the other side.

> *Comment: The lack of identity management and ubiquitous user/host authentication in the Internet impedes direct, secure user-to-user communication and hence promotes cloud-mediated (i.e., centralized) communication and service platforms. Such platforms typically do not support fine-granular, user-defined security policies, and this lacking can lead to overall more vulnerable systems. DDOS is a consequence of too many compromised devices. Mitigating DDoS by shielding servers in fortresses or via (centralized) CDNs is dealing with the symptom, not curing the disease.*

There seemed to be a shared pessimism among the workshop participants that the possibility of reversing this trend is not in a foreseeable future. *Why is security failing?* The participants offered several reasons:

1. Premature service releases ("rush to ship") often pay less attention to secure measures than they should, as security features are not attractive product differentiators that are visible to customers immediately. The cost of missing a deadline is immediate, but the cost of lacking security measures may only show up at much later time, and it is often the case that somebody else bears that cost.
2. As a technical community, we do care about well-designed security solutions. But getting security right is really hard, and interoperable security, which is needed to support decentralized systems and applications, is even harder. In contrast, it is much easier to develop closed, and hopefully less vulnerable, systems and applications.
3. The computer community at a whole is yet to be able to get to next level of software quality that can withstand attacks from well-resourced adversaries, such as the case of nation state sponsored attackers.

> *Comments: The above discussions lead to a few observations. First, 40 years of Internet development has trained a skilled network technical community who knows how to forward packets even in very large scale. The same network community is yet to gain expertise in building secure systems.*
> *Second, centralized vs. decentralized (federated) service designs, such as e-mail and micro-blogging etc., should be studied further with respect to the security and robustness implication of the respective design options. Security also needs to be qualified more precisely in this discussion. For example, federated e-mail may provide a more heterogeneous set of code bases and server deployments, but it also makes other threats, such as e-mail spam, more difficult to mitigate.*
> *Third, cryptographic authentication and encryption are well-understood and available security tools used in different protocols and systems today, such as in ubiquitous transport layer security realized by TCP and QUIC. However, the dominant use of these tools is supporting server authentications. In the absence of a general public key infrastructure for all networked entities, user (client) authentications still rely on vulnerable password-based mechanisms. Even for server authentications, the cryptographic key management is largely outsourced to third-party CAs, leaving security policy control outside the application domains to be secured.*

4.3 Looking Ahead

The discussion of future perspectives mainly focused on two aspects: i) Are we doomed? And ii) Can network protocols prevent centralization?

4.3.1 Are We Doomed?

As both panel speakers pointed out,

- personalized services require data, which leads to today's practice of service providers collecting all personal data at scale; and
- personal data has value, but only at scale. Today, individuals do not have means to monetize their own profile data, yet when this data is amassed in conjunction with the profile data from millions or billions of other users, the data collection becomes extremely valuable.

These observations paint a bleak picture of the future, indicating that it is unlikely for users to have control over their own data. If we consider data as valuable resources driving the digital economy, allowing application monopolies to have control over it presents a significant obstacle to achieving decentralization.

[4] https://dataprot.net/statistics/spam-statistics/

Comment: The key question is "who controls my data". Today it is the application providers, e.g., Google and Facebook. If one is to imagine an alternative solution for tomorrow, it could be the user oneself [10] – this could be done by storing the data either in user devices and/or in cloud storage in encrypted forms, which can be accessed with users' permissions to personalize their services. Such solutions are yet to be developed, with well defined, widely adopted user identities being the first requirement.

4.3.2 Can Protocol Designs Prevent/Mitigate Centralization?

A conclusion that one may derive from the discussion on the driving factors of Internet centralization is that network protocol designs *alone* may not be able to stop the market's movement towards centralization.

The Internet started as a decentralized network, where IP's distributed routing enabled a network of *decentralized connectivity*, and individual organizations set up their own servers for the few applications at the time (e.g., DNS, ftp, and e-mail), creating a decentralized world. In its early days, perhaps few people, if any, thought about the question of whether Internet would stay in that decentralized way.

Looking back, we note that protocols facilitate the movement of packets from one place to another, but do not restrict where this "other place" may be – as packet carriers, protocols do not dictate where packets go, it is applications that make the decision. We further note that, in the early days of Internet, many organizations ran application servers to provide services *for their own users*; they were not doing it as *revenue generating* business. Once network application services started being revenue-generating businesses, economy of scale drove them towards consolidation and concentration. Larger user groups make it more affordable for service providers to invest into service improvements, which then attract more individual users as well as organizations that are willing to outsource certain IT services and application given certain levels of affordability and quality. Challenges in mitigating ever growing network security threats add more attraction to outsourcing, or even make it necessary.

Comment: In the absence of effective regulations, it is the economy, not network protocols, that dictates the future direction of a system's evolution. It seems that the networking community did not recognize this fact early on, and the lack of this recognition led to complacency with the consolidating changes at their early stages.

5 SUMMARY

This report aims to provide a high-level summary of the workshop contributions and discussions at, and after, the workshop. All the workshop contributions can be found at [1], and the complete workshop recording is available at [11].

We note that Internet consolidation has been observed across all aspects of the Internet, ranging from network access to infrastructure and to application deployment. With limited time, however, the presentations and discussions at this workshop mostly focused on the aspect of users' data and application centralization.

We use this summary to capture the most important observations from the workshop. Reflecting on those observations, in this section we articulate the lessons to be learned and new insights to be derived. We hope that these lessons and insights can help aid the community's efforts in the centralization mitigation. These lessons and insights are from the authors' perspective, they are presented here to seek feedback from the community at large through open discussions. Our first goal is to reach a shared understanding on the problem space, before diving into the exploration of specific technical solutions.

5.1 Observations

Today's centralization is the result of unregulated economy of scale. This is the high-level answer to the question of how we got here, which is not affected by the technical specifics. Similar to the industrial revolution history described by Huston, the Internet revolution happened so quickly, the technical community seemed blinded on exactly where the train was heading to, and the regulatory sector fell behind. Consequently, the market has largely been left to run on its own to maximize large corporations' profits, which has run into conflicts with end users' interests in privacy and sovereignty.

We note that economies of scale motivate corporations to grow big, and that being big in size by itself is not the sole problem. Today's centralization concern is not about the size of those cyber giants, but rather, the control power and the influence they possess over users, and over society as a whole.

Importance of Security. We observed that security issues and corresponding threats are one of the major causes for unwanted centralization of infrastructure, in particular DDoS attacks whose power is far beyond individuals' protection ability. Currently, only centralized systems seem to have capabilities to absorb them.

Another security factor that drives centralization is the limitation of the existing Internet security framework. It is a web-server-focused security framework with the web PKI, which seems to have led to a server-biased communication style. Users in general do not have their own identities; by necessity they are assigned a unique ID by their application providers. Therefore, even local neighbors must go through centralized platforms and their authentication services in order to communicate.

An additional problem is that enabling secure communications (providing servers with valid certificate chains, implementing security building blocks correctly, etc.) is perceived (most likely correctly so) as complex and essentially un-manageable without expert knowledge and service-provider-level scale.

Decentralization via blockchain? The blockchain movement, often referred to as an enabler for *Permissionless Innovation*, has proposed to move the Internet out of centralization by replacing large parts of the Web and additional infrastructure with a new design based on anonymous blockchain technologies. In addition to several economic and technical issues that have been brought up by many people (see [12], [13] as two recent ones), we note that this blockchain movement attempts to apply a technical solution to mitigate a economical/political problem. As this report pointed out, economical forces have driven the Internet itself from a decentralized start into a consolidated state, a lesson that seems to deserve attention from the blockchain community.

5.2 Reflections

This workshop on centralization in the Internet helped deepen our understanding of the problem space. We recognize that technology alone is not responsible for the observed consolidation and centralization; rather, it is the uncontrolled economic force to blame. Early Internet applications were largely operated in a decentralized manner when they were provided as user services, and before adversaries recognized the value of this new cyberspace. Their move to revenue-generating business triggered the market force to drive further development towards consolidation, and security threats further intensified the move. Given centralization is fundamentally an economic problem, it cannot be mitigated away solely by technology solutions.

Society thrives on the balance between economy, regulation, and technology. Today, we see an imbalance which is tilted to economy, with the regulations facing challenges of figuring out exactly what should be regulated and how, and the technologies falling behind security threats. To effectively mitigate centralization, we need to hit the nail on the head. To that end, we view that effective regulation and legislation is a deciding factor in curtailing unconstrained market, and that the technical community holds the responsibility to inform the regulatory sectors of what/how to regulate, and work with them in concert to carry out new regulations effectively by providing new technical solutions that can curtail DDoS threats instead of merely absorbing them (by leveraging the power of centralized systems), and that can enable direct and secured user-to-user communications without reliance on cloud-platform provided authentication and policy enforcement services.

The topic of Internet centralization has captured attention in the IETF community over the last few years and inspired several ongoing efforts. As an example, draft-nottingham-avoiding-internet-centralization [14] enumerated the negative impacts of control centralization and suggested new technical solutions to mitigate them. We applaud all investigation efforts into new technical solutions. We also note that a key challenge to all new solutions is their wide adoption by the market. In the absence of effective regulations to nurture fair market competition, new solutions would not happen *automatically* if they focused on bringing benefits to users and society at large, but do not bring tangible benefits to the existing control parties. We hope that this report contributes to an open discussion that can help move the community to a shared understanding on the problem space, from there to derive effective next steps to progress towards decentralization.

Economic forces tend to perpetuate the continuous trend towards capital concentration and infrastructure centralization, and the lessons we have learned from Internet development – both technical and economical – should help us gain new vigilance to watch out future (re-)centralization, and to start mitigation efforts at its early stage by providing input to regulators and by adjusting technical solutions to meet the new challenges.

As our departing words: The Internet centralization problem will not solve itself. The networking community needs to take actions, now.

ACKNOWLEDGEMENT

We would like to thank all the workshop participants for their contributions, and thank Chris Riley for his valuable comments on an early version of the report. Dirk Kutscher was supported by the Guangdong Provincial Key Laboratory of Integrated Communications, Sensing and Computation for Ubiquitous Internet of Things and by the *Guangzhou Municipal Science and Technology Project 2023A03J0011*. Lixia Zhang is partially supported by the US National Science Foundation under awards 2019012 & 2019085. This paper reports the views of the authors, which do not necessarily represent the views of the funding agencies.

REFERENCES

[1] DIN Research Group. *DINRG Workshop on Centralization in the Internet: Workshop Material.* Online: https://datatracker.ietf.org/meeting/interim-2021-dinrg-01/session/dinrg, 2021. This web page includes Workshop Agenda, presentation slides, and all workshop submissions.

[2] Geoff Huston. *Is "Big" Necessarily "Bad"?* Online: https://datatracker.ietf.org/meeting/interim-2021-dinrg-01/materials/slides-interim-2021-dinrg-01-sessa-centrality-in-the-internet-geoff-huston-00, June 2021.

[3] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Online: https://www.amazon.com/Age-Surveillance-Capitalism-Future-Frontier/dp/1610395697, 2019.

[4] European Commission. *The General Data Protection Regulation (GDPR).* Online: https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en#fundamental-rights, 2018.

[5] Christian Huitema. *How Does Centralization Start?* Online: https://datatracker.ietf.org/meeting/interim-2021-dinrg-01/materials/slides-interim-2021-dinrg-01-sessa-christian-huitema-how-does-centralization-start-00, June 2021.

[6] Jari Arkko. *The influence of internet architecture on centralised versus distributed internet services.* Online: https://www.tandfonline.com/doi/pdf/10.1080/23738871.2020.1740753?needAccess=true, 2020. Published in JOURNAL OF CYBER POLICY.

[7] Loa Andersson, Elwyn B. Davies, and Lixia Zhang. Report from the IAB workshop on Unwanted Traffic March 9-10, 2006. RFC 4948, 2007.

[8] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou. Understanding the Mirai Botnet. USENIX, 2017.

[9] Rahaf Alkhadra, Joud Abuzaid, Mariam AlShammari, and Nazeeruddin Mohammad. Solar winds hack: In-depth analysis and countermeasures. In *2021 ICCCNT*, 2021.

[10] Steve Bellovin. *PRIVACY: MODERN CONCERNS.* Online: https://www.cs.columbia.edu/~smb/talks/ietf-privacy.pdf, 2019. IETF 105 Plenary Presentation.

[11] DINRG. *DINRG Workshop on Centralization in the Internet: video recording.* Online: https://youtu.be/1kbSbVjb1ZU, June 2021.

[12] David Rosenthal. *Can We Mitigate Cryptocurrencies' Externalities?* Online: https://blog.dshr.org/2022/02/ee380-talk.html, February 2022. Stanford Technical Seminar.

[13] Cory Doctorow. *Comprehensive synthesis of the technological, ecological and political critique of blockchainism.* Online: https://pluralistic.net/2022/02/14/externalities/#dshr, February 2022.

[14] Mark Nottingham. Centralization, decentralization, and internet standards. Internet-Draft draft-nottingham-avoiding-internet-centralization-05, IETF Secretariat, July 2022. https://www.ietf.org/archive/id/draft-nottingham-avoiding-internet-centralization-05.txt.