Electric Barons

Morlock Elloi

2018

Abstract

Modern technology of computing machines provides plausible disguise for the ideology of power. This ideology permeates the infrastructure and its design methodology, disarming opponents with pseudotechnological excuses. The two need to be separated.

In the Name of the Infrastrucure

If you live in San Francisco's Potrero Hill, there are several rational choices for getting to a Mission bar. Public transport - take 10 to SoMA, then 16 to Mission. Bicycle: down De Haro, then 17th. These choices frame your notion of "Mission bar". Going there is contingent on weather, available time and mood. Some Mission bars are out of consideration - too hard to park, too far away from transport, too much feces on the sidewalk. Some are rather convenient and thus become the natural choice. Over time, the inconvenient ones are forgotten. You will never go to Dovre Club.

This happened because MUNI made particular decisions on bus routes, SFMTA determined parking availability, the city planners laid down streets as they are now, and the city politics resulted in the current sidewalk feces distribution. Your natural bar choice was engineered by many contributors over long time. This is how cities are, and we got used to it. Enabling efficient traffic requires choices to be made - by politicians, bureaucrats, city utilities and services. One is free to use this infrastructure in many ways ... that it allows one to use it. It is impossible to step out of the infrastructure, and this is why having a say in its workings is important. On your way to the convenient bar you may consider all of this and perhaps table some thoughts for future actions.

Cities are shaped through constant clash between residents, real estate owners, investors, builders, politicians, parties and action groups. Anyone vaguely familiar with city politics knows how hard and slow it is to change anything. When a MUNI bus stop is too far away from your destination, or your street has potholes, or there are too many cars, you will not argue with the bus driver, pavement contractors or car drivers. You may contact your district supervisor to bring up the issue at next SFCTA/SFMTA meetings. First you need to find out that SFCTA and SFMTA exist, and what they do - it's all in the public record, and not so hard to figure out. Eventually your intervention may bear fruit and you will have a drink at Dovre Club. Or you may decide to move to some other place, with more agreeable politics and the infrastructure.

Different cities have different mentalities and have been influenced by different politics and development strategies. They all have one thing in common: city workings are observable. Changes in the city scene are obvious - construction sites, traffic jams, traffic enforcement, parking tickets, cops on the beat. The infrastructure is visible: streets in your neighborhood have bike lanes or they don't, they are congested or not, they are one way or two way. This visibility, in turn, means that one is in the position to make informed choices for the action.

What happens when another kind of traffic, communication between people, gets engineered by multiple interests? Our 'natural' ways to communicate are verbal, visual and touch, all physical and requiring proximity of the other person. Anything else requires some form of technology and infrastructure - from printed pages to fiber and satellite relays. Who are the builders, investors, real estate owners, politicians, action groups, utilities and services companies for this infrastructure? What does the road map look like? How does one initiate the change?

This is important to know, because the state and the layout of these roads may affect you far more than the city traffic jams. It may determine who you will know, who your friends and adversaries will be, what education and job you will have, who your future family will be, and how you will die. It is important to be able to see and recognize this infrastructure the same way the traffic jam or sidewalk feces are recognized, because one should not trust PR departments and experts claims that these things are or are not there. Effective politics and activism happen only after one spots traffic jams and feces.

But communication and data processing infrastructure in not visible, and politics and ideologies of its builders are far from obvious. There are many reasons for it, but this text is not about the history. It will try to show that politics and ideology of this infrastructure do exist, and how they affect people.

The ideology of the infrastructure goes deep and is often invisible to the involved actors. The participants generally believe that they are doing the best possible job. What is specific to engineering is that the governing ideology is often internalized as a technical issue, and is so presented to the insiders and the outsiders. The presumed difficulty to understand technicalities is used as a barrier to shield the ideology from the outsiders. The baffling part is that it also works on the inside. It is extremely hard to penetrate this construct and separate the ideology from the technology: the amount of its inherent nonsense can shame any belief system known to man. Yet it must be done.

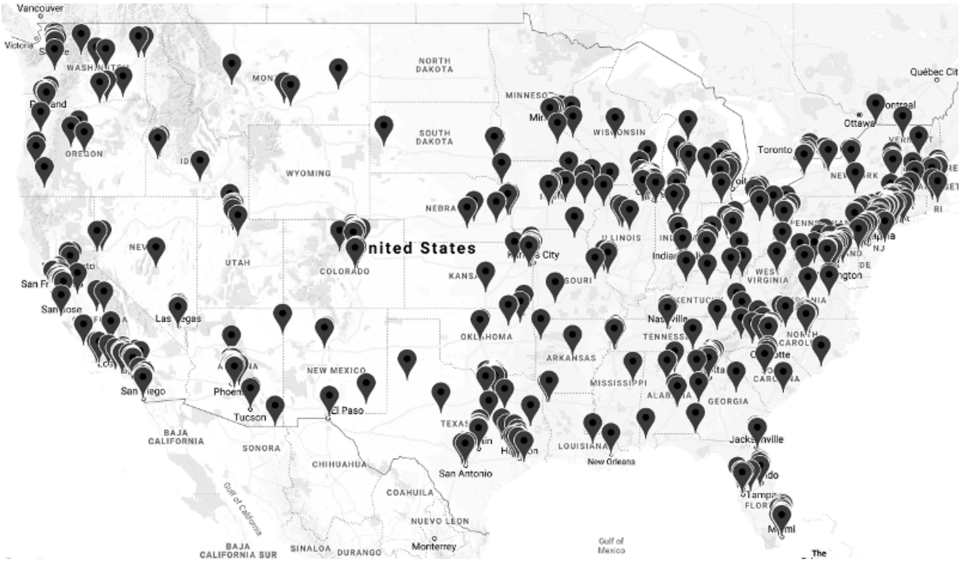
One aspect of this ideology is centralization. Centralization of traffic, directories, data bases, personal information, you name it. That center is somewhere where you are not. People interfacing machines built under this ideology are called 'users'. This is telling, as those, for example, driving machines or being driven by machines are called differently - drivers, pilots or passengers. The prospect of having millions obey machine instructions designed by the few is very seductive. In the previous times only select novelists, directors, songwriters and dictators enjoyed this replicative amplification of consequences, mostly for entertainment, enlightenment and indoctrination of the audience. Today the code determines conversations, money flows, employment, entertainment, and the rest of the life.

Perhaps the most sinister aspect is that it captures the energy of activism, which adopts the ideological canons and builds the same dystopian constructs, on the premise that they are now operated by the good guys, as if an Open Source cage is anything but a cage. The underlying fallacy, that the power will be used only for good purposes, becomes obvious always too late, when the energy and trust have been exhausted. Thus the useful idiots complete the ecosystem and seal it against the alternatives.

The infrastructural issues considered here are fundamental in nature, not about kinds of traffic on top of the infrastructure (social networks, search monopolies, etc.) The list is not exhaustive - only three examples scattered across huge domain. It is about the constrains the infrastructure imposes and inertia against the change that it creates, pretending to have technological justification, and about the need for a major redesign. It is obvious that this is a huge problem and uphill struggle, but it is a lesser problem than continuing down the current path. And finally, this is not a Luddite argument against the machines. We do need the machines, but designed and operated under a different ideology. What that ideology shall be we should determine through the established social institutions. The present we experience is not a technical problem.

The Server Tax

Servers are computers living in large numbers in buildings where space, air conditioning, power and bandwidth pipes are abundant. These are called 'server farms' or 'colocations' if servers are owned by multiple parties. Large enterprises own multiple farms. It's hard to estimate the total number of servers, but taking into account the 25 million annual server-class CPU chip sales, and using 4 years as a typical server life span, a round number of 100 million active servers is reached. This graphic from https://wiredre.com/us-data-center-list/ shows rough farm distribution in the US:



On the other hand, there are around 2.5 billion smartphones in the world, and at least as many other edge computing devices (PCs, tablets, smart TVs, IoT.) Taking into account that average server is 10-100 times more 'powerful' (in terms of CPU and storage) than a smartphone, it appears that the total edge computing power is likely greater than the total power of the servers, with trend favoring the edge. The 'edge' here means all computers in hands or homes of individuals, businesses, etc. The sheer weight of the edge silicon (metal from which chips are made) will be order of magnitude higher than the weight of the total server silicon, if it is not already so. Still, as there are several billion edge equipment owners, and only few million server owners, with less than few hundreds of the large ones, an server owners control computing power thousands and millions of times over that of the average edge equipment owner.

Computing power-wise, servers do not appear to be a dominant component of the whole system. Yet today servers control everything. Pretty much all edge devices talk directly only with servers. The centralization makes the traffic pattern look rather odd: as if residents of all 2612 San Francisco streets, when visiting another one, would all have to pass through the intersection of Market and Van Ness.

Almost every single 'app' and 'web site' is based on this paradigm. There is a server and there are thousands/millions/billions 'clients'. The underlying motivation is that the owner of the server has control over multitudes of clients. The edge equipment owners themselves do not need this centralization any more than San Francisco residents need to pass Market/Van Ness every time they go to the grocery store, but they have little choice. This is considered normal - from dreams of every startup that few will make and operate something that billions will use, to schools where engineers learn to make the servers work and to make the clients work.

Technical arguments for data storage concentration are weak. Today the entire Wikipedia can fit on a smartphone. Street maps of all places a person will visit in a lifetime are only few gigabytes in size. The amount of the 'new' content is relatively tiny. Yet servers insist on dishing out small crumbs of information from their centralized storage when it is required, enabling the server owners to know what edge devices are doing and when, almost like giving some change to a child to buy one ice cream from the store.

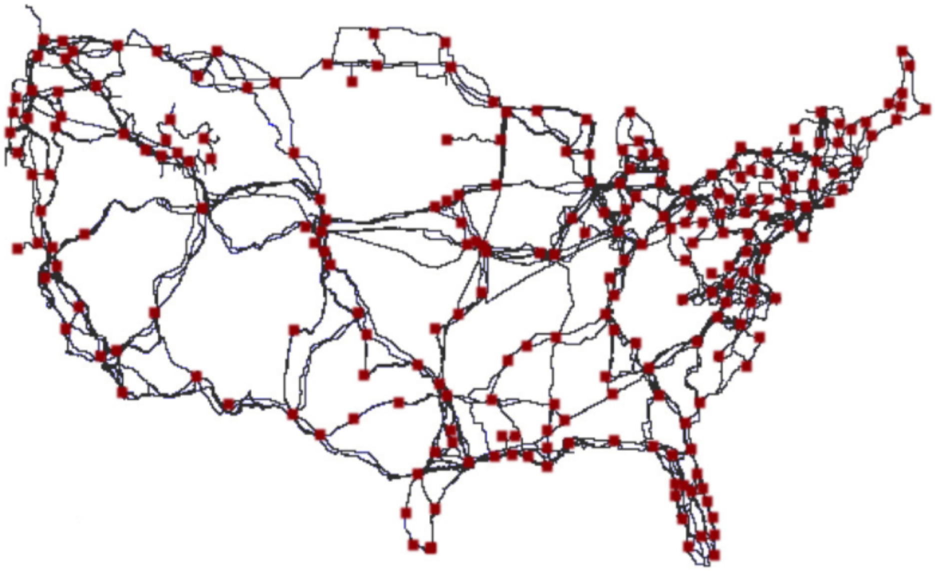
Can today's popular services exist with decentralized storage? The answer is yes, and there is empirical proof: as countries assert their sovereignty, the companies providing these services are compelled to move storage of data related to citizens to their respective countries, and make them subject to local laws. It doesn't seem that any of these services suffered because of this. What works on the country level will certainly work on any other level: city, municipality, household.

Technical arguments for computation centralization are even weaker. Edge devices do not benefit from it, for simple reason that, per device, there is far more power in the edge device itself than in the tiny fraction of the server apportioned to the device.

On the technical side there are solutions for distributed applications, naming, storage, resource discovery and source routing, so interventions against the server canon are political. The edge has to become its own center. It involves regulating against faraway processing and storage. There are parallels in the urban politics world: road builders never managed to make everyone go through the single toll ramp; many cities built barriers to chain stores; exploiting and parasitizing on human social and other instincts is customarily regulated by law: there are very few places in the world with industrialized prostitution. Political instruments already exist.

Cyberslum Concierge

The communications infrastructure consists of long-haul backbone of fiber lines, which in US more or less follow roadway infrastructure, and of Internet Service Providers (ISPs) that provide 'the last mile' connectivity between the backbone and individual participants. This is the layout of fiber in the continental US, from "InterTubes: A Study of the US Long-haul Fiber-optic Infrastructure, R. Durairajan et al 2015".



This looks more or less like the physical roadway infrastructure, where highways connect cities, each of which has its own street grid.

But there is a difference.

While you can get into your car in San Francisco and drive all the way to friend's apartment building in New York, enter the elevator and visit your friend, you can not do the same with information packet for your friend. The information packet will be stopped at the building entrance. The desk clerk would check if you friend expects the packet. If your friend did not alert the desk one minute or less before the expected packet arrival, your packet will be thrown out. In other words, you cannot send a surprise gift to your friend. The interesting question is how do you alert your friend? He cannot ask you, as your building has the same unsolicited packet policy. You two can never exchange packets directly. There have been numerous attempts to trick the entrance desk in letting unsolicited packets in, but they were never reliably successful, and thus never became basis for wide-spread direct connectivity.

Fortunately, there is another business nearby that will accept packets from anyone. In the machine world it's called a server. Both you and your friend can send packets to the server. When you send a packet to the server, you alert your building desk clerk that you expect something back from this particular server.

Your friend must do the same. In this way, the server acts as a meeting point to enable exchange between you and your friend. Needless to say, the server has to like both of you, and it is usually because it profits from each.

There are thousands times fewer servers than apartments, and the initial technical rationale for this entrance triage and server middlemen was that there are not enough street addresses to cover everyone, so only servers get to have public entrance for themselves, while everyone else shares the building address, and must deal with the desk clerk (by the way, the building street address itself changes, sometimes several times a day, so you cannot count on it being known.) The server addresses do not change. You have no choice but to use the server as middleman if you want to communicate with others. Except that there is a choice, a new style of addressing, available since 1998, called IPv6. It is not catching on, which may something to do with the lucrative server business. Even where IPv6 gets implemented, ISPs tend to cripple it or require, through contracts, that you will never receive unsolicited packets (in technical parlance, 'run a server'.) While you and your friend must use the intermediaries, server owners don't have this problem, so for example the New York Times is directly accessible, because it has its own servers.

This two-tiered system, on one side the inability to accept information directly and to have own permanent street address for ordinary people, and on the other side the privileged position for server operators, has deep influence both on edge participants and on the way edge computers are designed and used. The edge equipment owners acquire mentality of a homeless person, with no permanent address, and consider it normal. The edge computers must be tethered to various privileged servers, as they cannot communicate directly. This has shaped minds of engineers and frameworks and tools that they use and design. This is where centralized social networks and e-mail providers stem from. While this has not been a technical obstacle for 20 years now, it still defines the entire computing landscape.

The intervention is simple: lobby for unrestricted IPv6 and permanent addresses for everyone.

Celebrity-based Security

During WW2 Germans were using Enigma machine to encrypt world-wide military radio traffic, including the one between the central command and submarines. Allies managed to break the Enigma encryption in 1941, but they didn't tell Germans. Instead, they allowed some ships to be sank, and in other cases the scout airplanes would 'accidentally' spot the submarine whose position was known from decrypted traffic, so that Germans would not suspect broken cipher.

Germans eventually found out about this in 1973, and had hard time believing it. Many texts have been written on the topic of hubris of Enigma designers. This 3-decade gap between successful cryptanalysis and public learning about it is typical in cryptography, and should be accounted for when designing security methods. It does not make cryptography useless - the cryptanalysis will not be used against low value targets, and some ships can be sunk.

In general, the cryptanalysis (breaking a cipher) is as hard if not harder than the cipher design itself. It took Allies considerable effort - many man years - to perform this cryptanalysis feat. One obvious way to stay ahead of the cryptanalytic curve is to keep introducing new ciphers. In WW2 the machines executing both the encryption and the cryptanalysis were electromechanical and expensive to construct, so introducing new ciphers was a very slow process; yet Germans modified their Enigma machine three times in 10 years, once in the middle of WW2.

One would think that today, when all cryptography is done in software that can run on any computer, this staying ahead of the curve principle is a norm. It is not. On the contrary, the ideology of cryptography preaches the exact opposite: never do custom cryptography, always use the standard one, approved by the experts who know better than you. Like in every ideology, there is truth in this: designing strong ciphers and strong security systems is very hard. But it completely misses the balance of power relationships and threat models. It is also obvious that the cipher designer hubris is still cryptographic constant.

While it is hard to estimate how many 'licensed' experts participate in design of ciphers, it is possible to estimate the upper boundary: International Association for Cryptologic Research has about 2,000 members. On the other hand, the unverified lower estimate for the number of mathematicians working for just one state agency (NSA) is about 10,000, so we can assume that world-wide there are at least 10 times more brains paid to cryptanalyse than engaged in designing ciphers, and this ratio could easily be one hundred times. It gets worse: the total number of block ciphers widely used today ('approved by experts') is four (AES, Camellia, ARIA, ChaCha20-Poly1305, all published between 1998 and 2007), and the total number of key exchange protocols deemed secure and widely used is also four (RSA, DH, EC, GOST, all published between 1976 and 2005.) Compromise of either key exchange or block cipher compromises the whole system. At this point, tens of thousands of cryptanalysts had ten years to compromise four algorithms, designed by less than dozen experts. The official mantra to use these four is part of the power equation mandating uniformity of the protective gear, and is permeating both academia and the industry. On the other side, no diplomatic service, military or government communications appear to use any of these: "Serious countries (USA, UK, Germany, France) do not use foreign algorithms for high-security needs" (Eric Filiol at ESIEA.)

Here again we see the 'few for many' principle rearing its ugly head, facilitating centralized control and related compromises.

Should everyone design their own ciphers, millions of companies and individuals designing their own terribly weak ciphers, new one every year? There is no automated way to cryptanalyse even naively weak ciphers (and many would make not so naive ones.) Tens of thousands of cryptanalysts cannot begin to chip away even at the weak security of millions of new custom and unpublished ciphers every year. First they would have to figure out what is the cipher, and then break it. It takes time, even if it is a variant of ROT-13.

What would happen is leveling of the playing field, engaging brains against brains, on the scale that cryptanalysis cannot keep up with. The targeted cryptanalytic approach would still work, but with limited amount of targets, and will likely be no different in overall amount of breaches than today's successful targeted hacking of computer systems. The mass snooping would cease to exist.

The compliance with the cryptographic ideology which forbids custom ciphers boggles the mind, as it prevents the only hope for changing the power imbalance: recruitment of the raw brain power.

The intervention is straightforward: as different applications do unique things and have unique code, they should have unique ciphers, and change them often. The probability that your system will be broken into by targeted hacking will remain roughly the same, and the probability that your system is continuously open book for the major adversaries goes down to zero.

What Can Be Done?

The luxury of not caring for the infrastructure and leaving it to the experts and industries involved must be abandoned. Efforts on superstructure levels (information monopolies, copyrights, data ownership and collection, freedom of speech. etc.) are pointless when the infrastructure embodies and petrifies diametrically opposite models. No amount of smoke and mirrors and assurances that the operators are honorable and law-abiding will ever change that. This infrastructure was successfully removed from the public view and it's time to do hard work and start looking at it.