Anais do VI Simpósio Internacional LAVITS: “Assimetrias e (In)Visibilidades: Vigilância, Gênero e Raça”

Coletando dados sobre o Capitalismo de Vigilância nas instituições públicas do ensino superior do Brasil

LEONARDO RIBEIRO DA CRUZ[1]
FILIPE DE OLIVEIRA SARAIVA[2]
TEL AMIEL[3]

Resumo
Nos últimos anos, houve um crescimento na oferta de plataformas educacionais às instituições públicas de ensino e de pesquisa no Brasil por grandes empresas do setor, particularmente as empresas Google e Microsoft. Acordos baseados na oferta de serviços gratuitos à instituições e entes públicos tem em grande parte, como contrapartidas, a coleta, tratamento, utilização e comercialização de dados comportamentais de seus usuários. Os acordos revelam uma grande assimetria em relação ao vetor da coleta de dados. As empresas têm, potencialmente, acesso a uma grande quantidade de dados das instituições públicas. Esse artigo tem como objetivo apresentar a metodologia e os resultados encontrados na investigação da seguinte questão: quais instituições públicas de ensino superior no Brasil fizeram acordo com as empresas Google ou Microsoft para gerenciamento dos seus sistemas de tecnologia de informação? Problematizamos a relação entre grandes empresas e sistemas de gerenciamento de ensino e comunicação das universidades públicas. Apresentamos dados primários atuais sobre o tema gerados por um script que será disponibilizado como software livre. Coletamos dados visando identificar servidores de e-mail das instituições, que sabemos (por estudo prévios) ser um forte indicativo de parcerias. Apresentamos os dados da investigação de 104 instituições estaduais e federais públicas brasileiras em formato aberto e com visualização georreferenciada, discutindo suas consequências para o ensino público, vigilantismo e soberania nacional associada à produção científica do país.

Palavras-chave: Capitalismo de Vigilância; Softwares Educacionais; Google Education; Microsoft 365.

Introdução

Nos últimos anos, temos acompanhado, como pesquisadores, o crescimento na oferta de plataformas educacionais às instituições públicas de ensino e de pesquisa no Brasil por grandes empresas do setor. acordos ou modelos de negócio baseados na oferta de serviços gratuitos a instituições e entes públicos têm em grande parte, como contrapartidas, a coleta, tratamento, utilização e comercialização de dados comportamentais de seus usuários (PARRA, CRUZ, AMIEL & MACHADO, 2018).

Contudo, existe uma grande assimetria em relação ao vetor da coleta de dados. Por um lado, as empresas têm, potencialmente, acesso a uma grande quantidade de dados das instituições públicas -

1 Professor e Pesquisador, doutor em Sociologia pela Universidade Estadual de Campinas (UNICAMP), professor adjunto da Faculdade de Ciências Sociais da Universidade Federal do Pará. leocruz@ufpa.br.

2 Professor e pesquisador, doutor em Engenharia Elétrica pela Universidade de São Paulo, professor adjunto da Faculdade de Computação da Universidade Federal do Pará. filipesaraiva@ufpa.br.

3 Professor e pesquisador, doutor em Instructional Technology pela University of Georgia, professor adjunto do Departamento de Métodos e Técnicas da Faculdade de Educação da Universidade de Brasília. amiel@unb.br.

desde dados pessoais de alunos, professores e funcionários, dados comportamentais dos usuários dos aplicativos educacionais e dos dispositivos (tablet, celular, computador), dados de rendimento escolar dos alunos e professores e até dados de comunicação institucional e de pesquisa. Por outro lado, como usuários e pesquisadores com interesse em dimensionar o mercado de dados, temos dificuldade tanto em saber se nossas informações estão protegidas quanto em obter informações oficiais sobre os termos, os acordos e os reais parâmetros dos dados coletados

Esse artigo tem como objetivo apresentar a justificativa, a metodologia e os resultados encontrados na investigação da seguinte questão: quantas instituições públicas de ensino superior no Brasil fizeram acordo com a Google ou a Microsoft para gerenciamento dos seus sistemas de tecnologia de informação e quais são elas? Não há dados oficialmente divulgados pelas empresas nem pelas instituições sobre o assunto, salvo notícias publicadas pelas assessorias de imprensa de algumas universidades públicas. A partir do desenvolvimento de um *script* que será disponibilizado como software livre, coletamos dados visando identificar servidores de e-mail das instituições públicas de ensino superior alocados nas máquinas das empresas pesquisas, que sabemos (por estudo prévios) ser um indicativo de parcerias. Apresentamos os dados da investigação de instituições estaduais e federais brasileiras, discutindo suas consequências para o ensino público, vigilantismo e soberania nacional associada à produção científica do país.

Capitalismo de vigilância e assimetrias de informação.

Nos últimos anos, temos visto o crescimento e a hegemonização de um modelo de negócio assente na captura, tratamento e comercialização de dados dos usuários da internet. Os sites que visitamos, os e-mails que recebemos, os assuntos que curtimos, nossa geolocalização e até o que conversamos em frente a dispositivos de comunicação digital viram dados que alimentam um mercado cuja finalidade é a predição e o controle do comportamento dos usuários da rede. Quanto mais atividades conduzimos e quando mais conteúdo publicamos nessas plataformas e serviços das grandes empresas desse mercado (notadamente as empresas Google, Amazon, Facebook, Apple e Microsoft, conhecidas como GAFAM), mais nossa sociabilidade se transforma em dados comercializáveis e mais imperceptível se torna a modulação dos ambientes da rede, que conduzem nosso comportamento e as nossas escolhas (ROUVOIR & BERNS, 2017).

Para apresentar esse mercado, usamos o conceito “capitalismo de vigilância”, cunhado pela economista alemã Shoshana Zuboff no artigo “Big Other: capitalismo de vigilância e perspectivas para uma civilização da informação” (2018). Nele, a autora tem o objetivo de apresentar uma novo impulso de acumulação capitalista realizado através da exploração, tratamento e comercialização de dados. Com foco na atuação da empresa Google, considerada por ela a pioneira dessa nova lógica de

acumulação, Zuboff disserta sobre uma nova forma de capitalismo de informação que "procura prever e modificar o comportamento humano como meio de produzir receitas e controle de mercado" (ZUBOFF, 2018: 18).

> A Google, até agora, triunfou no mundo em rede através da construção pioneira dessa nova forma de mercado, que é uma variante radicalmente descolada e extravagante do capitalismo de informação, que identifico como capitalismo de vigilância. Rapidamente se tornou o modelo-padrão de negócios na maioria das empresas e startups, em que as rotineiras estimativas de valor dependem de "olhos", mais do que de receita, para prever a remuneração dos ativos de vigilância (ZUBOFF, 2018).

Utilizamos o conceito de capitalismo de vigilância pois é, até agora, o que melhor consegue delinear os pressupostos econômicos de um tipo de acumulação que consegue transformar dados de comportamento em mercadoria. . Contudo, o desenho do conceito desenvolvido por Zuboff não é suficiente para compreendermos o capitalismo de vigilância como um sistema global e assimétrico, pautado nas desigualdades existentes no próprio desenvolvimento do sistema capitalista neoliberal.

Evgeny Morozov, em uma resenha crítica ao novo livro da Zuboff, afirma que o maior problema do desenvolvimento do conceito de capitalismo de vigilância é que a autora não conseguiu apresentar o que há de "capitalismo" no conceito. Segundo ele, o livro da autora perdeu a oportunidade de avançar conceitualmente ao não apresentar a relação dele com a lógica de desenvolvimento do capitalismo neoliberal, principalmente em relação a sua atuação assimétrica em diferentes partes do globo. Para o autor,

> há uma razão para acreditar que, da mesma forma que o capitalismo industrial tinha um padrão específico de exploração em determinados territórios e populações - um para os centros de poder e consumo, outro para a periferia e a produção - o mesmo acontece com o capitalismo de vigilância (MOROZOV, 2019).

Rafael Evangelista, em uma resenha sobre o mesmo livro, chegou a conclusões parecidas. Segundo ele, falta ao conceito uma perspectiva que não universalize as relações ocorridas nas principais democracias liberais, mas que consiga compreender como esse novo modo de acumulação se insere nas assimetrias de poder globais, em especial às relacionadas à divisão internacional do trabalho e do conhecimento (EVANGELISTA, 2019). Contudo, o autor não invalida a potencialidade do conceito. Pelo contrário, ele acha possível atualizá-lo a partir de novas perspectivas "que deem conta de descrever um funcionamento que evidencie a possível persistência e aprofundamento de assimetrias norte-sul"(Evangelista, 2017: 250).

Contudo, independente da perspectiva ou da conceitualização teórica, quem se dispõe a pesquisar empiricamente a atuação das grandes empresas do mercado de dados se depara com um problema em comum: a dificuldade de coletar dados ou conseguir informaçoes objetivas sobre a atuação local delas. Com nossa experiência de pesquisa sobre o pacote educacional da Google no Brasil, podemos

dizer que é muito difícil encontrar informações sobre a atuação da Google nas escolas e universidades públicas. A empresa divulga poucas informações e não responde a pedidos de entrevista. Grande parte de seus acordos com as instituições públicas não ocorre por vias contratuais e não há processos licitatórios, não deixando rastros que possam ser acessíveis via Lei de Acesso à Informação[4]

Mesmo as informações disponíveis são muito pouco objetivas e não nos permite ter clareza do funcionamento desse mercado. A pesquisadora Marta Kanashiro, por exemplo, ao investigar as políticas de privacidade do acordo entre a Universidade Estadual de Campinas e a Google, nomeou o conjunto de documentos, políticas de uso e termos de privacidade de "boneca matrioska", em que um documento faz referência a outro, que faz referência a um terceiro (Kanashiro, 2016). Nós também já escrevemos sobre a dificuldade de compreender as políticas de privacidade da Google e como isso se assemelha a uma estratégia de ofuscamento de sua atuação enquanto ator em um mercado (PARRA, CRUZ, AMIEL, MACHADO, 2018: 80-81).

Maria Lindh e Jan Nolin, ao analisarem as informações relativas à privacidade dos acordos do *Google Suite for Education* nas escolas públicas da Suécia no artigo *Information we colect: surveillance and privacy in the implementation of Google Apps for Education* (2016), chegaram à conclusão que a falta de clareza sobre a atuação da Google faz parte de uma série de artifícios retóricos para dificultar ou impedir o acesso ao conhecimento dos procedimentos realizados pela empresa. Segundo eles, é criada uma "zona cinzenta" de atuação em que não conseguimos ter um conhecimento assertivo sobre seu funcionamento.

Essa ocultação do modo de funcionamento das empresas do capitalismo de vigilância nas tecnologias educacionais está presente em quase todas as partes do mercado. Um relatório de pesquisa da Electronic Frontier Foundation intitulado "*Spying on Students: School-Issued Devices and Student Privacy*" (ALIM, CARDOZO, GEBHART, GULLO & KALIA, 2017) e que investigou a utilização dos softwares educacionais utilizados em escolas públicas estadunidense aponta que a falta de transparência nas relações com as empresas é a principal reclamação de professores, pais, alunos e administradores escolares. A segunda é que essa falta de transparência transfere para pais e alunos a responsabilidade de buscar informações e de se protegerem de eventuais inseguranças no uso dos aplicativos.

4 ver exemplo de resposta via LAI sobre o caso em http://www.consultaesic.cgu.gov.br/busca/dados/Lists/Pedido/Item/displayifs.aspx?List=0c839f31%2D47d7%2D4485%2Dab65%2Dab0cee9cf8fe&ID=579060&Web=88cc5f44%2D8cfe%2D4964%2D8ff4%2D376b5ebb3bef e http://www.consultaesic.cgu.gov.br/busca/dados/Lists/Pedido/Item/displayifs.aspx?List=0c839f31%2D47d7%2D4485%2Dab65%2Dab0cee9cf8fe&ID=658577&Web=88cc5f44%2D8cfe%2D4964%2D8ff4%2D376b5ebb3bef].

O que nos deparamos é com a grande assimetria de informações em nossas relações com as empresas do capitalismo de vigilância – se por um lado não podemos ter clareza sobre seu modo de atuação na qual estamos amplamente implicados, do outro, temos nossas ações, enquanto usuários de plataformas e serviços digitais, constantemente rastreada.

Frank Pasquale (2015) denomina essa lógica assimétrica de "Black Box Society". As empresas do mercado de informação, cujo modelo de negócio se baseia na extração crescente de dados de rastreamento, se utilizam de critérios e brechas técnicas e legais para impor uma lógica de segredo sobre seu modo de atuação. Essa lógica da ocultação nos impede de entender, investigar, regular e participar conscientemente do mercado. Por outro lado, o modo de atuação dessas empresas se baseia na coleta e análise incessante de uma grande quantidade de dados sobre nossa atuação. Essa posição unilinear da privacidade acaba por impor uma nova lógica de autoridade sobre os indivíduos e suas ações. Por fim, a utilização dessa lógica sobre a gestão dos dados captados e a reutilização automática desses dados nos algorítimos que compõem as plataformas e serviços acabam por influenciar nossas futuras decisões sem que consigamos nos informar e atuar de forma autônoma nesses espaços.

Shoshana Zuboff (2018), mais uma vez, nos apresenta algumas chaves para que possamos compreender a dinâmica entre privacidade e segredo na lógica de acumulação do capitalismo de vigilância. Segundo ela, as empresas desse mercado conseguem, através da ampliação em grande escala da utilização de suas ferramentas, concentrar os direitos de privacidade e o consequente poder de escolher o que deve ou não ser mantido em sigilo enquanto esse direito é retirado de seus usuários. A concentração desse direito mantém a obscuridade das operações de vigilância e cria novas assimetrias de conhecimento e poder.

> [...] a Google sabe muito mais sobre sua população de usuários do que estes sabem sobre si mesmos. De fato, não há meios pelos quais as populações possam atravessar essa divisão, dados os obstáculos materiais, intelectuais e proprietários necessários para a análise de dados e a ausência de feedback loops. Outra assimetria assenta no fato de que o usuário típico tem pouco ou nenhum conhecimento sobre as operações comerciais da Google, sobre a ampla gama de dados pessoais com que contribui para os servidores da Google ou sobre a retenção desses dados ou, ainda, como eles são instrumentalizados e monetizados. Já é bem sabido que os usuários têm poucas opções significativas para a autogestão de privacidade. O capitalismo de vigilância prospera na ignorância do público (ZUBOFF, 2018: 50).

Como pesquisadores, por um lado, precisamos compreender a atuação dessa lógica de acumulação em nosso cotidiano. Ela está ocorrendo em grande escala e transformando nossa realidade de uma forma acelerada. Por outro, temos muita dificuldade para qualificar essa atuação e mensurar o avanço dessa lógica sobre as outras. Em nossas pesquisas sobre a relação entre as empresas do capitalismo de vigilância e instituições públicas de ensino superior do país – que envolve, dentre

outras coisas, a transferência de todos os dados dos e-mails institucionais para servidores das empresas, geralmente alocados fora do país – não conseguíamos dados para responder questões como: quantas instituições haviam feito acordos com a Google ou a Microsoft? Com que velocidade esses acordos estão sendo realizados? Qual é a dimensão da quantidade de dados institucionais que estão sobre controle das empresas?

Hackeando a Google e a Microsoft: Um *script* para mensurar os acordos

Para contornar a falta de transparência e impossibilidade de acesso a base de dados de instituições universitárias públicas brasileiras que aderiram aos programas *Google for Education* e *Microsoft Office 365 for Schools & Students*, desenvolvemos nessa pesquisa um script disponibilizado como software livre chamado "get-mx-universities.py" que automatiza a consulta e tratamento de dados sobre as respostas enviadas por servidores de domínio quando consultados sobre quais computadores são responsáveis por gerenciar o tráfego de e-mail para um domínio em específico. Essa técnica foi aplicada no universo de endereços de e-mail de 104 universidades públicas brasileiras, tanto federais quanto estaduais.

Essa seção se dedicará a explicar o funcionamento do *script* e os conceitos relacionados com tráfego na Internet que permitiram esse tipo de pesquisa. Cabe destacar que a explicação aqui buscará ser didática, evitando expôr conceitos mais complexos sobre redes de computadores. O intuito também é permitir um melhor entendimento e possibilitar o reuso do script para pesquisas desse tipo em outros cenários.

IP e DNS: Arquitetura de Provimento de Serviços na Internet

Computadores são acessíveis na internet a partir de um endereço numérico que especifica a "localização" da máquina na rede. O protocolo que especifica o endereçamento das máquinas é chamado de Internet Protocol (IP), e é um dos pilares do funcionamento da internet (Kurose et al, 2007). Abaixo, temos 2 exemplos de endereços IPs, respectivamente para a versão 4 e versão 6 (IPv4 e IPv6) do protocolo:

IPv4: 200.144.248.41

IPv6: 2001:12d0:c000:91::41

O endereçamento numérico, apesar de usual em termos técnicos e permitir a organização da internet nos seus diversos serviços disponíveis, é pouco adequado para a utilização convencional dos usuários: ele requereria a memorização de diferentes valores numéricos para acessarmos os vários serviços que utilizamos diariamente.

Nesse contexto, existe o protocolo de serviço de nomes de domínio - Domain Name Services (DNS) - que traduz um endereço textual, chamado domínio, em numérico, indicando para a aplicação o endereço do computador que provê o serviço que se está buscando (Kurose et al, 2007). Uma analogia bastante comum em livros e artigos sobre DNS comparam-no com uma lista telefônica, que informa o número de telefone relacionado com a pessoa que se quer contatar.

A consulta ao DNS é, em geral, a primeira operação realizada por qualquer aplicação que utilize algum serviço na internet (Tanembaum, 2003). Ao se digitar em um navegador web o domínio “google.com” e pressionar o “Enter”, o navegador irá fazer uma consulta a servidores DNS na internet, que procurarão o endereço IP relacionado com o endereço “google.com” em consulta. Após encontrar, o DNS retornará para o navegador o valor, que no momento da escrita desse artigo é “172.217.162.174”. Esse endereço permitirá ao navegador acessar o computador correto que proverá o serviço requisitado (no caso, como se trata de um navegador web, o próprio site do Google).

Para o exemplo da Figura 1, quando o navegador web no computador pessoal tenta acessar o domínio “usp.br”, ele primeiramente realizará uma consulta ao servidor DNS, que no caso retornará o endereço IPv4 relacionado com aquele domínio - portanto, o endereço numérico que indicará qual computador é responsável por prover os sites “usp.br”.

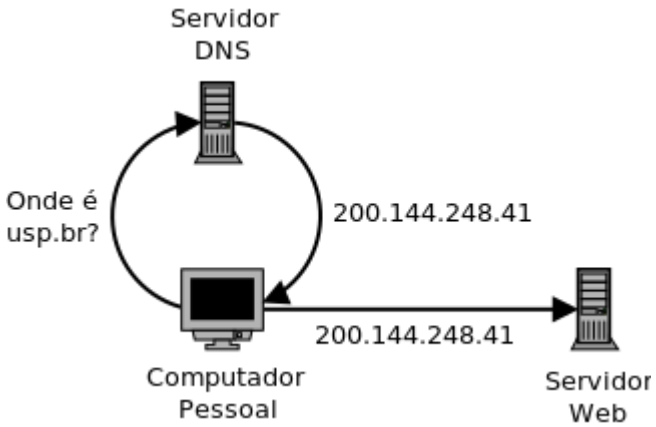


Figura 1

consulta DNS para acessar o site “usp.br”

É importante destacar que o servidor DNS também pode redirecionar endereços textuais para outros endereços textuais, o que acaba demandando da aplicação uma nova consulta até que se consiga o endereço numérico do computador que provê o serviço buscado.

Outra característica que cabe comentar sobre arquiteturas de serviços na internet e como acessá-los, é que um mesmo computador pode hospedar diferentes serviços ao mesmo tempo: não é incomum

que uma mesma máquina disponibilize na rede aplicações como sites (portanto, um servidor de aplicações web), e-mail (servidor de e-mail), banco de dados, e outros.

Também muito comum é que diferentes computadores possam responder por um mesmo domínio, mas disponibilizando serviços diferentes: assim, quando uma aplicação web faz uma consulta DNS, o endereço respondido será o computador que hospeda o site (serviço web) para aquela requisição; entretanto, se a aplicação que faz a consulta é um cliente de e-mail, o DNS respondido será o endereço do computador que serve o serviço de e-mail para aquele domínio - que pode ser o mesmo responsável pelo serviço web ou outro computador.

Dessa forma, retornando ao exemplo dessa seção, quando um navegador web tenta acessar o domínio “usp.br”, tem como valor do DNS retornado o computador que hospeda o serviço de sites da universidade; já quando utilizada uma aplicação de e-mail, por exemplo, enviando um e-mail para algum endereço “@usp.br”, o computador responsável por receber a mensagem será aquele que gerencia o servidor de e-mails da instituição. No momento em que esse artigo é escrito, o computador para essa última tarefa está disponível no endereço “ASPMX.L.GOOGLE.COM”, com réplicas acessíveis também em “ALT1.ASPMX.L.GOOGLE.COM”, “ALT2.ASPMX.L.GOOGLE.COM”, “ALT3.ASPMX.L.GOOGLE.COM” e “ALT4.ASPMX.L.GOOGLE.COM”.

A Figura 2 apresenta, de maneira gráfica, essa situação. Uma consulta ao DNS para o endereço web “usp.br” retornará o endereço numérico “200.144.248.41”, que aponta para um determiado computador; já a consulta para um endereço de e-mail “@usp.br” retornará o endereço textual “ASPMX.L.GOOGLE.COM”, que aponta para um outro computador, diferente do anterior. Em verdade, a resposta textual dessa última demandaria outra consulta ao DNS para se obter o endereço numérico, entretanto, em prol da concisão, essa outra consulta não é apresentada na figura.

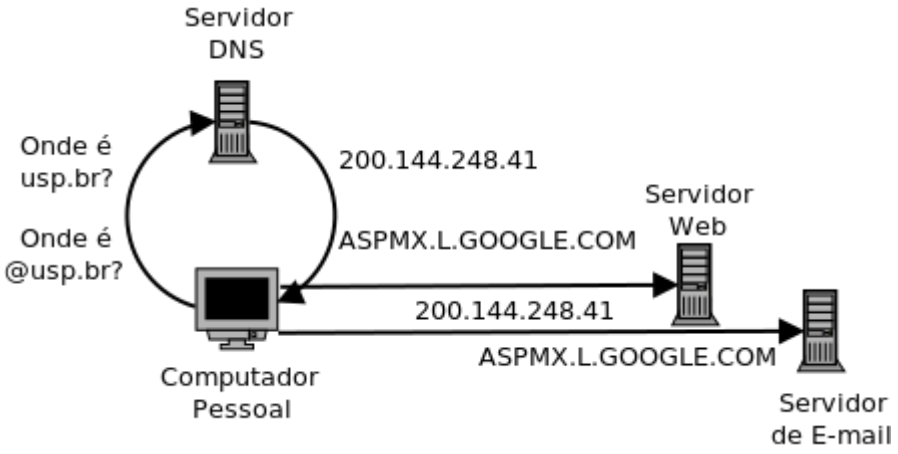


Figura 2

Duas consultas ao servidor DNS, uma para web e outra para e-mail, tem respostas diferentes que encaminham para computadores diferentes

Esse mecanismo, portanto, permite que servidores de e-mail de universidades utilizem o domínio/endereço de seus respectivos sites nas contas dos seus usuários (por exemplo, “reitoria@usp.br”), apesar do gerenciamento do serviço em si ser realizado em um servidor externo à instituição, sob controle de uma empresa como as que estamos pesquisando nesse artigo: Google ou Microsoft.

Descobrindo o Dono do Servidor de E-mail

O momento em que se define quais computadores serão responsáveis por quais serviços atrelados a um domínio se dá na configuração de DNS desse domínio (Costa, 2006). Nessa etapa, o responsável pelo domínio indica os endereços dos computadores responsáveis pelo servidor web, de e-mail, e demais serviços que forem providos sob aquele domínio.

Nas configurações do DNS, o técnico responsável pelo domínio especifica determinadas opções que apontam o endereço do computador que proverá cada serviço. Nessas opções, tem-se um código sobre o “tipo de serviço”, que tem valores textuais padrões para cada tipo de serviço: por exemplo, “A” para endereços IPv4, “AAAA” para endereços IPv6, “NS” para outros servidores DNS, “CERT” para certificados, entre outros.

O tipo de serviço que especifica o gerenciador de e-mails para o domínio é identificado pela sigla “MX”. Na Figura 3, temos um exemplo com diferentes “types” (tipos de serviço) atrelados ao domínio guides.ovh. Nela, a coluna Target especifica os computadores que serão redirecionados para atender cada serviço. Nesse exemplo, o servidor que responderá pelas trocas de e-mails do domínio será o mx1.mail.ovh.net.

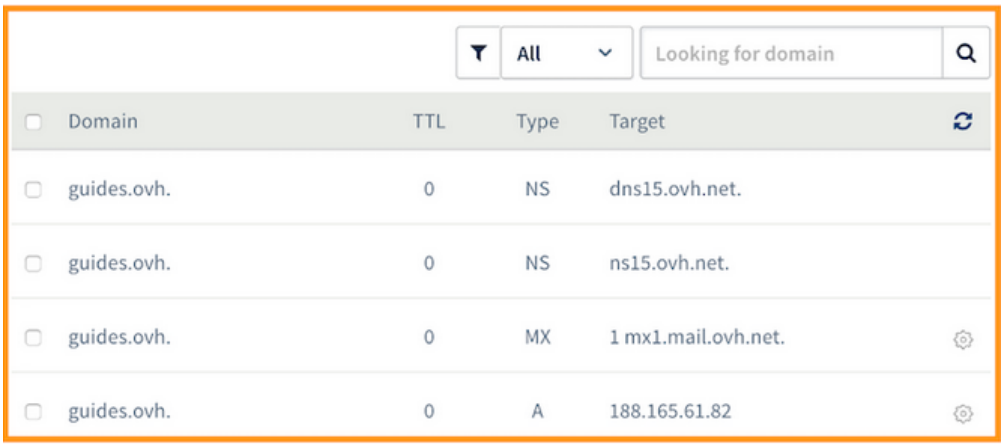


Figura 3

Configuração DNS para serviços diferentes do domínio guides.ovh (Fonte: https://docs.ovh.com/gb/en/domains/web_hosting_how_to_edit_my_dns_zone/)

É possível saber qual computador é responsável pelo gerenciamento de e-mails de um domínio a partir da verificação do redirecionamento dado ao MX na configuração de DNS do domínio. Para realizar essa atividade, uma empresa americana relacionada com o fornecimento de tecnologias bases para uso na internet, a Internet Systems Consortium (https://www.isc.org/), disponibiliza para o sistema operacional GNU/Linux um programa chamado “host”, que faz a verificação do redirecionamento de diferentes serviços de um domínio.

Para executar esse programa, em sua forma padrão, basta utilizar como argumento um domínio específico. Por exemplo, executando o comando para o domínio utilizado pela USP teremos a seguinte saída:

$ host usp.br

usp.br has address 200.144.248.41

usp.br has IPv6 address 2001:12d0:c000:91::41

usp.br mail is handled by 5 ALT1.ASPMX.L.GOOGLE.COM.

usp.br mail is handled by 10 ALT3.ASPMX.L.GOOGLE.COM.

usp.br mail is handled by 1 ASPMX.L.GOOGLE.COM.

usp.br mail is handled by 5 ALT2.ASPMX.L.GOOGLE.COM.

usp.br mail is handled by 10 ALT4.ASPMX.L.GOOGLE.COM

Na primeira linha temos a execução do “host” utilizando “usp.br” como entrada. As demais linhas são as respostas dadas pelo comando.

As duas linhas seguintes informam, respectivamente, os endereços IPv4 e IPv6 do servidor web relativo ao domínio. As linhas restantes indicam quais são os computadores responsáveis pelos e-mails da USP - perceba que são 5, todos eles relacionados ao Google. Essa quantidade de computadores diz respeito ao balanceamento de carga do serviço, onde há disponibilidade de vários computadores para garantir que o serviço de e-mail atenda à demanda da instituição.

Portanto, para identificar qual computador gerencia um e-mail relacionado com um domínio, devemos executar o comando “host” sobre aquele domínio e buscar as linhas que identificam os responsáveis pela funcionalidade “mail” - o MX na configuração do DNS.

Assim, o que o script desenvolvido realiza é a execução automatizada do comando “host” sobre uma lista de endereços de domínios de universidades brasileiras. Por realizar essa operação de forma massiva, o script recebe como entrada um arquivo texto convencional com a lista dos domínios que ele irá pesquisar, um em cada linha.

Ao iniciar a consulta dos domínios informados no arquivo, o script filtrará as respostas para capturar apenas uma saída relacionada com o servidor de e-mail, descartando as demais.

Capturada a saída, o script faz uma série de comparações no endereço do computador responsável por prover o serviço de e-mail, buscando nele subcadeias de caracteres comuns utilizadas pelas empresas Google e Microsoft para endereçarem seus computadores na internet. As configurações de subcadeias utilizadas foram as seguintes:

Google: google e gmail;

Microsoft: microsoft e outlook.

O script é extensível o suficiente para que novas subcadeias sejam adicionadas ou mesmo novos fornecedores sejam inseridos, sem prejuízo de performance para a execução da funcionalidade.

Enquanto o script executa as consultas, caso ele encontre na resposta do servidor algum dos padrões elencados nas subcadeias, ele incrementará um número que indica a quantidade de universidades que utilizam servidores de e-mail naquele fornecedor. Ao final, o script informará a quantidade de universidades que utilizam os servidores em cada um dos fornecedores, os respectivos percentuais, e o total de domínios analisados. Essas saídas são impressas no próprio terminal durante a execução do script mas também são salvas em um arquivo texto convencional com o log da execução realizada.

O script desenvolvido chama-se “get-mx-universities” e é escrito em Python 3, fazendo uso apenas de bibliotecas padrões dessa linguagem de programação. Os testes realizados foram feitos

exclusivamente no Linux e requer que o programa “host” esteja instalado na máquina.

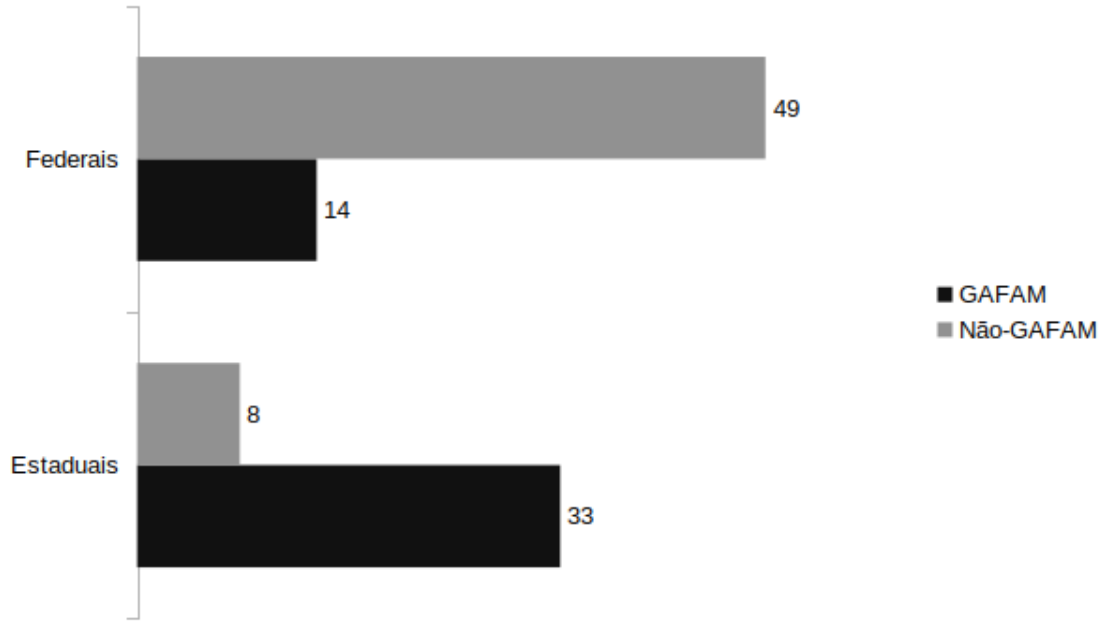
O “get-mx-universities” está disponível sob a licença permissiva MIT, possibilitando usos para os mais diversos fins e das mais diferentes maneiras, sem afetar qualquer outro programa que porventura venha a reutilizá-lo, seja este livre ou proprietário. O código-fonte também está disponibilizado em um repositório no Gitlab (https://gitlab.com/ccsl-ufpa/get-mx-universities/) e preservado no repositório Zenodo (DOI: 10.5281/zenodo.3249155).

Para essa pesquisa, foram coletados domínios de endereços de e-mail de 64 universidades federais e 41 universidades estaduais, totalizando 104? endereços de e-mail. Os endereços utilizados estão disponíveis no repositório do *script*.

Mapeamento

Com base nos dados coletados, georreferenciamos de forma automática e manual o posicionamento aproximado das universidades públicas federais e estaduais. Geramos um arquivo CSV (arquivo em texto separado por vírgulas que alimenta um mapa baseado no OpenStreetMaps (http://bit.do/gafam). Uma vez atualizados os dados, a visualização reflete as mudanças. Em fase posterior o script será atualizo para automaticamente gerar um CSV e automatizará a publicação periódica de dados. No entanto, do mesmo mapa já é possível baixar os dados de forma aberta (licença CC-BY).

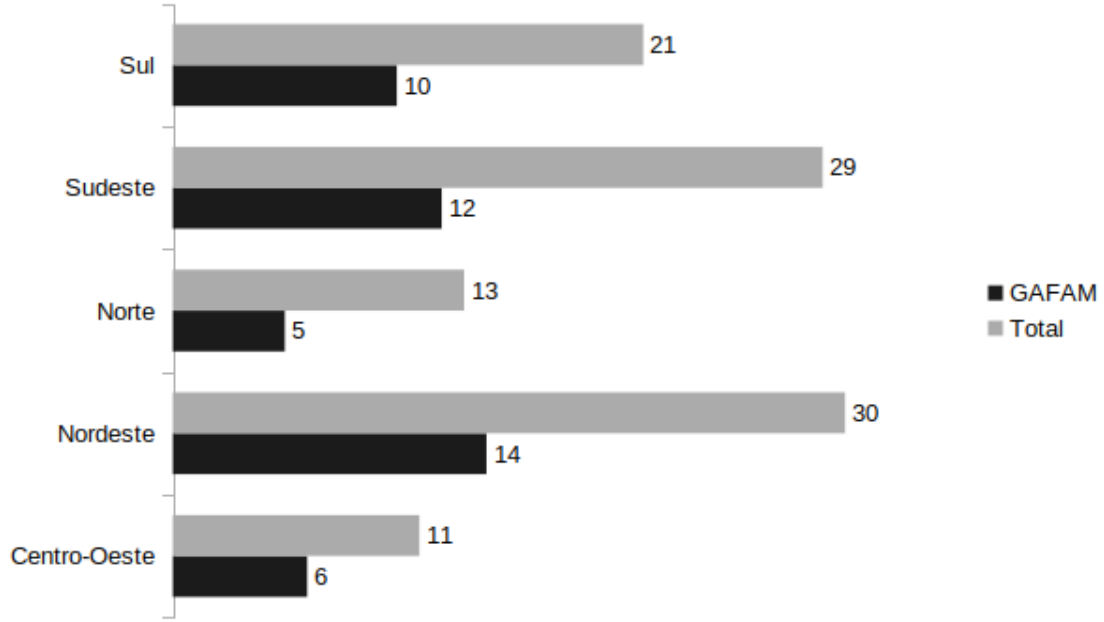
Resultados



Os dados indicam que 80% das universidades estaduais investigadas foram demarcadas como tendo

servidores MX associadas ao Google ou Microsoft. Os caso se inverte nas federais, onde o número chega a 22%. No total, temos indicação que 47 das 104 instituições se enquadram no parâmetro de análise. Importante anotar que das federais, somente uma instituição (UnB) e dentre as estaduais, somente 6 tem seus servidores MX associados à Microsoft, sendo o restante associado ao Google.

A distribuição por região geográfica apresenta um cenário equitativo de presença da universidades associados ao Google e Microsoft.



O mapa nos permite observar a distribuição geográfica de forma a perceber a presença das empresas de forma equânime em território nacional. No mapa os pontos vermelhos indicam as instituições que foram identificadas como tendo servidores MX associados à Google ou Microsoft. Os pontos e *clusters* em verde indicam universidades sem essa identificação, o que aponta para servidores ou serviços próprios.

Análise

Conseguimos constar, na nossa coleta de dados, uma evolução na migração para plataformas de vigilância. Podemos atribuir essa migração para serviços privados internacionais à diversos fatores. Alguns desses fatores foram levantados em análises detalhadas da transição de algumas instituições (Parra, Cruz, Amiel, Jorge, 2018) e incluem a redução de custos, o sucateamento das áreas de tecnologia da informação nas instituições, o reconhecimento do uso recorrente dessas ferramentas por docentes, servidores e discentes (nesse caso, a institucionalização da parceria normaliza uma prática vigente) e como já indicamos inúmeras vezes, um subdimensionamento radical das consequências éticas, econômicas e políticas dessa transição.

Observamos um quadro de redução e incerteza orçamentária nas universidades, que tornaram manchete e foram alvo de protestos ao longo do ano de 2019. A imprevisibilidade orçamentária faz com que acordos com empresas sejam extremamente atrativos, já que garantem um serviço confiável e de qualidade a custo orçamentário baixo. Ignora-se, é claro, o custo de implementação e constante interação técnica com as empresas, que de certo é diminuto se comparado com o curso de manutenção de sistemas.

Para além de um agravamento do quadro, motivado por congelamentos, cortes e "contingenciamentos" de verbas destinadas às universidades públicas, a revogação do Decreto nº 8.135, de 4 de novembro de 2013[5] contribuiu para o quadro migratório. Esse explicitava em seu primeiro Artigo:

> As comunicações de dados da administração pública federal direta, autárquica e fundacional deverão ser realizadas por redes de telecomunicações e serviços de tecnologia da informação fornecidos por órgãos ou entidades da administração pública federal, incluindo empresas públicas e sociedades de economia mista da União e suas subsidiárias.

Deixava claro que serviços de correio eletrônico, bem como o processamento e armazenamento de dados deveriam seguir a mesma norma. Este foi revogado pelo então presidente interino Michel Temer através do Decreto 9637, de 26 de dezembro 2018, que não propõe orientações específicas sobre a conduta das instituições de ensino superior, mas prega a "integração e cooperação entre o Poder Público, o setor empresarial, a sociedade e as instituições acadêmicas" (Art 3, inciso XV). Ao remover a necessidade de manutenção de estruturas de armazenamento e processamento de dados em instituições públicas federais, o decreto permite que cada vez mais instituições federais de ensino insiram seus alunos, professores e pesquisadores, de forma compulsória, no modelo de negócio do capitalismo de vigilância.

Conclusões

Essa comunicação é animada pela necessidade e dificuldade de conseguir dados sobre o modo de atuação das empresas do capitalismo de vigilância no Brasil. A escassez de informações sobre esse mercado dificulta que possamos atuar nele, regular sua movimentação e nos tornarmos conscientes de sua conduta em nossa vida cotidiana.

Esse ocultamento das atividades faz parte de uma estratégia de controle de mercado. A falta de regulação e o desconhecimento sobre a atuação das empresas impedem a concorrência ao mesmo tempo em que dificultam a discussão acerca da decisão de incluir milhares de pessoas, de forma

[5] Esse decreto foi promulgado pela presidenta Dilma Rousseff como uma forma de defesa das comunicações de dados da administração pública federal após as denúncias de espionagem realizadas por Edward Snowden.

compulsória, em seus canais de coleta de dados, inclusive todas as suas comunicações institucionais.

Esse estudo contribui para discutir a possibilidade de abrir a caixa preta do capitalismo de vigilância e trazer à tona dados que embasam o fenômeno ainda pouco estudado de forma empírica - a transição da gestão de tecnologia da informação das universidades e redes públicas de educação para grandes empresas de mídia multinacionais. O *script* criado permite que consigamos rastrear o movimento das empresas, levantar hipóteses sobre sua atuação e discutir seus efeitos.

Os próximos passos desta pesquisa é coletar periodicamente dados sobre os servidores de e-mail das universidades públicas brasileiras para acompanhar a realização dos acordos. Além disso, alimentaremos o script para incluir dados de instituições públicas de ensino de outras localidades para, enfim, compará-las. Com isso, pretendemos obter dados sobre a atuação global do capitalismo de vigilância sobre tecnologias educacionais.

Referências

ALIM, F. CARDOZO, NATE. GEBHART, GENE. GULLO, KAREN & KALIA, AMUL (2017). **Spying on Students: School-issued devices and student privacy**. Electronic Frontier Foundation. Disponível em <https://www.eff.org/issues/student-privacy>

COSTA, D. G. (2006). **DNS-Um Guia para Administradores de Redes**. Brasport.

EVANGELISTA, Rafael. 2018. **Review of Zuboff's The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power**. Surveillance & Society 17(1/2): 246-251.http://library.queensu.ca/ojs/index.php/surveillance-and-society/index|

KUROSE, J. F., ROSS, K. W., & ZUCCHI, W. L. (2007). **Redes de Computadores e a Internet: uma abordagem top-down**. Pearson Addison Wesley.

LINDH, Maria; NOLIN, Jan (2016). **Information We Collect: Surveillance and Privacy in the Implementation of Google Apps for Education**. European Educational Research Journal,Vol 15, Issue 6, pp. 644 – 663.

PARRA, H.Z.M.; Cruz, L.; AMIEL, T.; MACHADO, J. (2018). **Infraestruturas, Economia e Política Informacional: o Caso do Google Suite For Education**. MEDIAÇÕES, v. 23 n. 1, p. 63-99, Jan./un. 2018. Disponível em: http://www.uel.br/revistas/uel/index.php/mediacoes/article/view/32320/pdfAcesso em 3 de agosto de 2018.

PASQUALE, Frank (2015). **The black box society**. The secret algorithms that control money and information.Harvard University Press.

TANENBAUM, A. S., & WETHERALL, D. (2003). **Redes de Computadores.** Editora Campus.

Zuboff, Shoshana (2018). **Big Other: capitalismo de vigilância e perspectivas para uma civilização de**

informação. IN BRUNO, Fernanda *et al*.Tecnopolíticas da Vigilância: perspectivas da margem. Boitempo.