Resumos do Workshop de Medições da RNP

Disciplina: IC0060 - Internet do Futuro
Docente: Leobino Nascimento Sampaio
Discente: Lucas Mascarenhas Almeida

PALESTRA 1: ESTUDO DE TÉCNICAS DE APRENDIZADO PROFUNDO PARA REPRODUÇÃO DE DADOS DE MONITORAMENTO DE REDES COM GARANTIAS DE ANONIMIZAÇÃO

VINICIUS MOTA (UFES). PROJETOS PMON (CHAIR: ALEX MOURA) - 10:30 - 11:00HS.

A palestra consiste na apresentação de um projeto do professor Vinicius Mota (UFES) e grupo. Com o avanço da socialização das tecnologias do *Big Data*, grandes quantidades de dados sensíveis passaram a trafegar diariamente na internet, nestes cenários questões como privacidade e segurança de dados tem sido debatidas cada vez mais. No Brasil desde 2020 tem vigorado a Lei de Proteção a Dados (LGPD) que dentre outras técnicas, indica a anonimização e/ou pseudo-anonimização para que banco de dados com dados pessoais e/ou sensíveis se tornem adequados para manipulação. É a partir desse contexto que o professor Vinicius e grupo identificaram uma oportunidade de aplicação de *Deep Learning* para reprodução de dados de monitoramento de redes garantindo a devida anonimização.

Foram apresentadas duas principais alternativas que vigoram hoje quando o tema é uso de dados de monitoramento de redes. A primeira e talvez a mais popular é o uso de dados sintéticos para simulação de redes, a partir de, por exemplo, distribuição de probabilidade normal, cauda longa e etc, é possível simular o comportamento da rede. A principal desvantagem dessa abordagem é que os dados sintéticos não conseguem simular tão bem a realidade, devido a dificuldade de parametrização e limitações dos modelos de distribuição, já a grande vantagem é por gerar dados intrinsecamente anônimos, convergindo com as recomendações da LGPD. A segunda alternativa apresentada é justamente a utilização de dados reais, e essa apesar de ser muito interessante do ponto de vista de fidedignidade dos dados, obviamente esbarra na LGPD e na não trivialidade de anonimização, além disso compromete a reprodutibilidade do experimento já que dados reais vão estar reproduzindo um contexto de espaço e tempo muito particular.

Foi a partir desses cenários que surgiu a proposição de uma nova alternativa, a aplicação de *Generative Adversarial Network* (GAN's) para produção de dados sintéticos. A expectativa é que essa abordagem consiga capturar de forma mais fidedigna (do que a abordagem de simulação tradicional) as características dos dados reais sem deixar de manter a anonimização dos dados, permitindo, dentre outras coisas, a reprodutibilidade dos experimentos. Essa alternativa ainda está em fase de validação, um primeiro resultado apresentado foi que as GAN's, técnica inicialmente aplicada para processamento de imagem, são adequadas também para predição de séries temporais. Outro resultado importante foi que os modelos conseguem simular a rede de forma muito mais fidedigna do que os modelos tradicionais, essa performance depende do tamanho da amostra, mas com menos de 1 hora de amostragem (apenas 20 ou 30 minutos) os resultados já são bem satisfatórios.

Apesar dos resultados promissores, algumas hipóteses precisam ser testadas e desafios ainda precisam ser superados. Um primeiro desafio é a capacidade de processamento (ou a falta dela), com uma amostra de minutos o apresentador destacou que as vezes passava dezenas de horas para processar todos os dados e gerar resultados. Outra questão é a sintonia da anonimização, se o modelo mimetizar exatamente o comportamento da rede a anonimização ficará comprometida, podendo ficar vulnerável a ações de reversão, além de também impactar na reprodutibilidade dos experimentos. Isto é, ainda não está claro qual é o patamar que o modelo não pode ultrapassar para manter os dados anonimizados.

Esse tipo de projeto vai ao encontro dos debates em torno do tema Internet do Futuro, mais especificamente a abordagem evolucionária. A garantia da anonimização dos dados contribui com a mitigação de um problema relevante quando se trata de Internet, a privacidade e segurança dos dados (inclusive na abordagem *clean-slate* essa questão também é relevante). Outra importante contribuição é que uma vez que as hipóteses estejam validadas e um *framework* de simulação que garanta anonimização, reprodutibilidade e performance esteja disponível, haverá uma ampliação significativa no horizonte de novas aplicações e tecnologias para redes. Isso permitirá, por exemplo, a criação de modelos de predição de falha para *backbone* da rede, o que, dentre outros benefícios, aumentaria a segurança e performance da Internet.

PALESTRA 2: TELEMETRIA ADAPTATIVA USANDO APRENDIZADO POR REFORÇO PROFUNDO EM REDES DEFINIDAS POR SOFTWARE

SIDNEY LUCENA (UNIRIO). INICIATIVAS NACIONAIS (CHAIR: ALEX VIEIRA) - 14:30 - 15:00HS.

A palestra consiste na apresentação de um artigo publicado no Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC) por Debora H. Job (UNIRIO), Sidney C. de Lucena (UNIRIO) e Pedro Nuno Moura (UNIRIO), sendo professor Sidney o responsável pela apresentação no Workshop de Medições da RNP. Com o aumento da popularização de tecnologias evolucionárias como a redes definidas por software (*Software Defined Network* - SDN), pesquisas e projetos passaram a ampliar as possibilidades de integrações, uma que vem sendo explorada é o uso de telemetria *in-band*. Apesar de ser uma combinação muito interessante existem alguns desafios, um dos principais é o custo associado. Foi pensando nisso que o professor Sidney e grupo estão propondo uma telemetria adaptativa que usa aprendizado por reforço em SDN para, dentre outras coisas, reduzir custos.

É fato que a SDN habilitou um avanço significativo na programabilidade das redes. Com a possibilidade de customização, ajuste e reprogramação usando soluções SDN os monitoramentos especializados de rede se tornaram cada vez mais sofisticados. Atrelado a isso oportunidades de aumentar a granularidade desse monitoramento de rede vem sendo fomentadas a partir do uso de telemetria *in-band*, esse tipo de técnica consiste basicamente em alocar medidas da rede dentro de pacotes de dados específicos, o que combinado com a programabilidade, permite que metadados sejam inseridos no *header* do pacote. Esse tipo de sistema entrega informações importantes sobre a rede, permitindo ampliar os horizontes na área de monitoramento de redes, dentre outras aplicações é possível ainda combinar

com os algoritmos de inteligência artificial, formando o que vem sendo chamado pela comunidade científica de *Knowledge-defined networking* (KDN).

Apesar das diversas vantagens que esse tipo de sistema pode oferecer na área de monitoramento de redes os custos associados ainda são muito altos, o que pode acabar tornando financeiramente inviável. Não é difícil notar que quanto mais informação trafega na rede, maior será o custo associado, e pensando em um sistema que combina telemetria, SDN, IA e monitoramento a quantidade de dados pode escalar de forma indesejada, não só do ponto de vista de tráfego de rede, mas também no armazenamento desses dados.

A partir desse cenário Sidney e equipe estão propondo usar aprendizado por reforço em uma rede neural profunda para que em um contexto de telemetria *in-band* e SDN seja possível regular a intensidade da telemetria, o modelo aprenderia o comportamento da rede (na rede) usando aprendizado por reforço profundo para decidir quando aumentar ou diminuir a telemetria. Para isso, o grupo está desenvolvendo um framework denominado SmartMON que combina algoritmos de aprendizado por reforço profundo e o contexto de telemetria em SDN. Os principais cenários de testes apresentados são fundamentados na combinação de perfis de uso de tráfego e ajustes de hiperparâmetros dos modelos, os resultados são promissores e realmente atendem as expectativas de redução de custo sem perda de desempenho do modelo.

Apesar dos bons resultados é importante observar que a amostra coletada representa um universo muito particular, o que implica na necessidade de tentar generalizar com novos testes para novos cenários. Outro ponto é que existe uma dificuldade de acesso a uma infraestrutura real para colocar o sistema em produção, apesar de ser uma ambição do grupo chegar nesse patamar de experimentação, durante a apresentação foi mencionado que hoje não está no alcance deles e que estão abertos para colaborações nesse sentido.

Essa iniciativa é um exemplo explícito de abordagem evolucionária, mostrando não só as vantagens de como a internet atual pode ser resiliente, mas também as várias desvantagens e dificuldades associadas ao modelo vigente. A SDN é uma tecnologia que vem sendo explorada, a partir desse movimento novas aplicações que antes eram inviáveis, ou até impensáveis, estão sendo propostas. Mesmo com várias limitações (como o IP) essa tecnologia vem permitindo avanços significativos em relação a questões fundamentais da Internet como redução de custos.