MAC6958
Tópicos Avançados em Ciência de Dados para Redes de Computadores

- **Prof. Daniel Macêdo Batista**
  Prof. Roberto Hirata
- **DCC-IME-USP**

- 19/3/2020

Roteiro

- **Detecção de anomalias em segurança de redes** (Kétly Gonçalves Machado)
- **Redes de datacenters** (Lucas Batista Gabriel)
- **Estrutura da Internet** (Caio Lente)

Detecção de anomalias em segurança de redes

- Anomalia pode ser considerada como um valor “fora do normal”. Há várias formas de calcular. Uma delas é por meio de outliers
  - Um outlier em um conjunto de valores é um valor fora do intervalo:

    $[Q_1\text{-}k(Q_3\text{-}Q_1), Q_3\text{+}k(Q_3\text{-}Q_1)]$

    - K = 1,5
    - Q1 = Primeiro quartil: valor abaixo do qual 25% de todos os valores se encontram
    - Q3 = Terceiro quartil: valor abaixo do qual 75% de todos os valores se encontram

Detecção de anomalias em segurança de redes

- No caso de fluxos em redes de computadores pode ser difícil resumir todas as informações em um único número
  - É comum ter várias grandezas sendo medidas (tamanho médio dos pacotes, intervalo entre pacotes, endereços IP – que podem estar associados a latitude/longitude da geolocalização, protocolo, etc..)
  - A anomalia pode ser encontrada observando quantas das grandezas para um fluxo estão em outliers e alertar se x% delas são outliers. O x depende do quão rígido o administrador de sistemas é (x menor → mais rígido. x maior → mais rígido). O x afeta falsos positivos/negativos

Detecção de anomalias em segurança de redes

- Muitas vezes uma anomalia não necessariamente é algo ruim, como no caso de flash crowds (ou efeito slashdot) em que muitas pessoas estão de fato acessando um serviço, de forma legítima, ao mesmo tempo
- Não é fácil distinguir flash crowds de DDoS
  - Flash crowd → deve ser tratado alocando mais VMs para dar conta dos acessos
  - DDoS → o tráfego deve ser bloqueado

Detecção de anomalias em segurança de redes

- Medir a rede para detectar anomalias pode ser muito custoso computacionalmente pois os pacotes muitas vezes só fazem sentido quando analisados em conjunto (importante analisar o fluxo de dados – data stream - e não os pacotes individuais)
  - Análise do fluxo de rede: capturar os pacotes espelhando uma porta de um switch por exemplo com scapy/libpcap/Wireshark (Muitos dados mas vê tudo)
  - Análise dos logs de serviços de rede: normalizar e processar logs de servidor web, BD, etc... com syslog/rlog (Menos dados e mais interpretação mas pode perder alguma informação)

Detecção de anomalias em segurança de redes

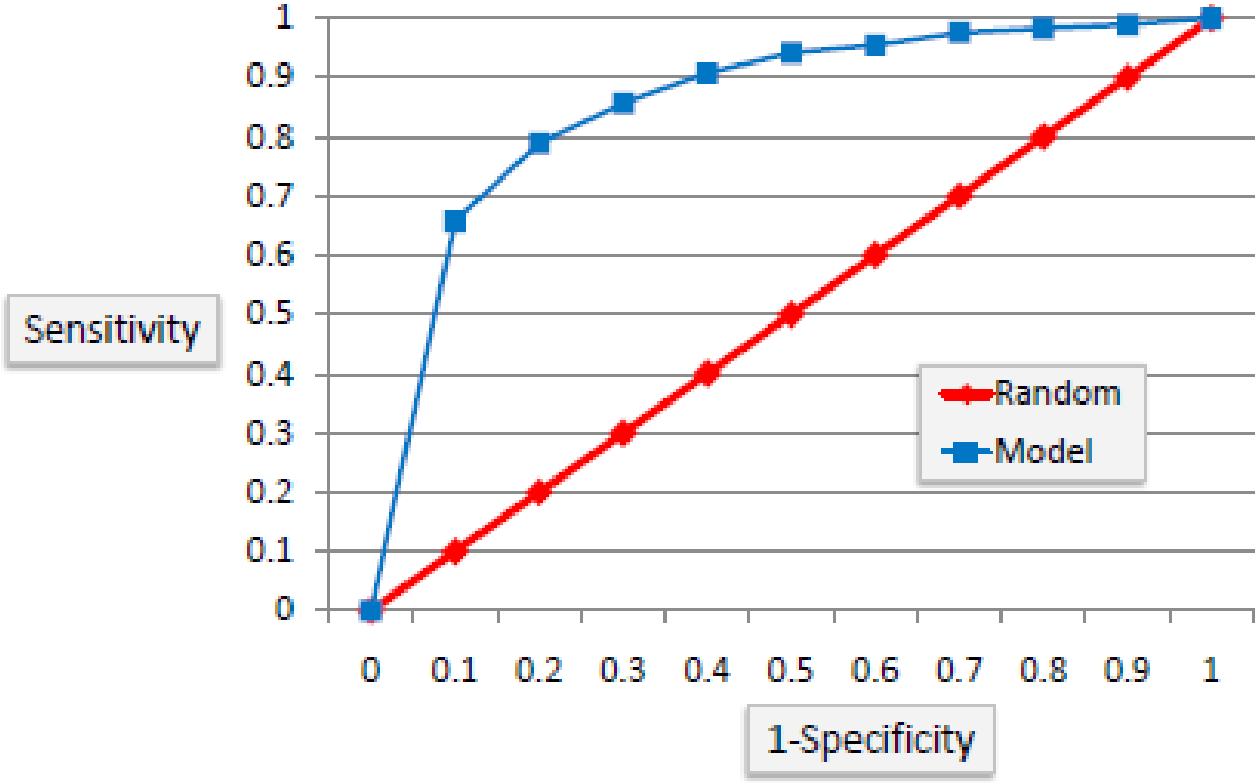
- Medir a rede para detectar anomalias pode ser muito custoso computacionalmente pois os pacotes muitas vezes só fazem sentido quando analisados em conjunto (importante analisar o fluxo de dados – data stream - e não os pacotes individuais)
  - Ser humano muitas vezes tem que fazer parte do processo
  - Distribuir as medições com sensores espalhados pela Internet permite antecipar os efeitos negativos de uma anomalia (empresas de segurança fazem isso e oferecem como serviço)

Detecção de anomalias em segurança de redes

- Importante mostrar quão bom é um sistema de detecção de anomalias
  - Velocidade da detecção
  - Falsos positivos e falsos negativos
    - Matriz de confusão
    - AUC – Area Under the Curve

•AUC: Area Under the Curve

- Se área < 0,5 → Pior que o aleatório. Tem que inverter os rótulos



Redes de datacenters

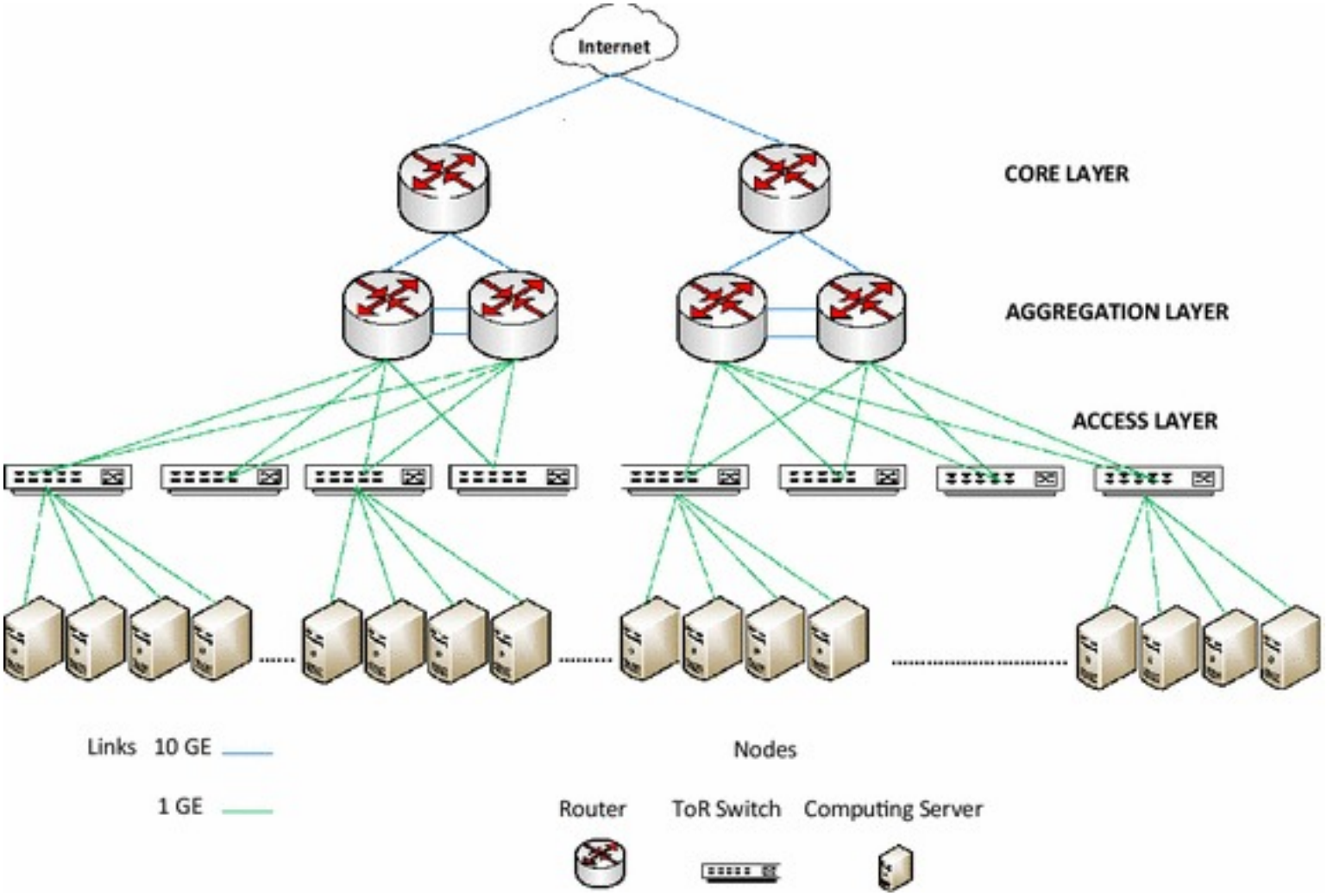
- Datacenters (Antes eram comum o termo CPD – Centro de Processamento de dados) precisam de topologias de rede eficientes
- Muitas máquinas se comunicando para resolver um problema complexo (computação em nuvem, computação em grade, etc...)
- Fisicamente as máquinas precisam estar próximas entre si e entre os elementos de interconexão (roteadores e switches)

Redes de datacenters

- Topologias mal planejadas podem causar:
  - Alta temperatura → Que levam a falhas do hardware e a alto consumo de energia tanto pelas máquinas quanto pela necessidade de muita refrigeração
  - Alto atraso → Que prejudicam a QoS
  - Mais cabeamento → Que aumentam o custo financeiro

Redes de datacenters

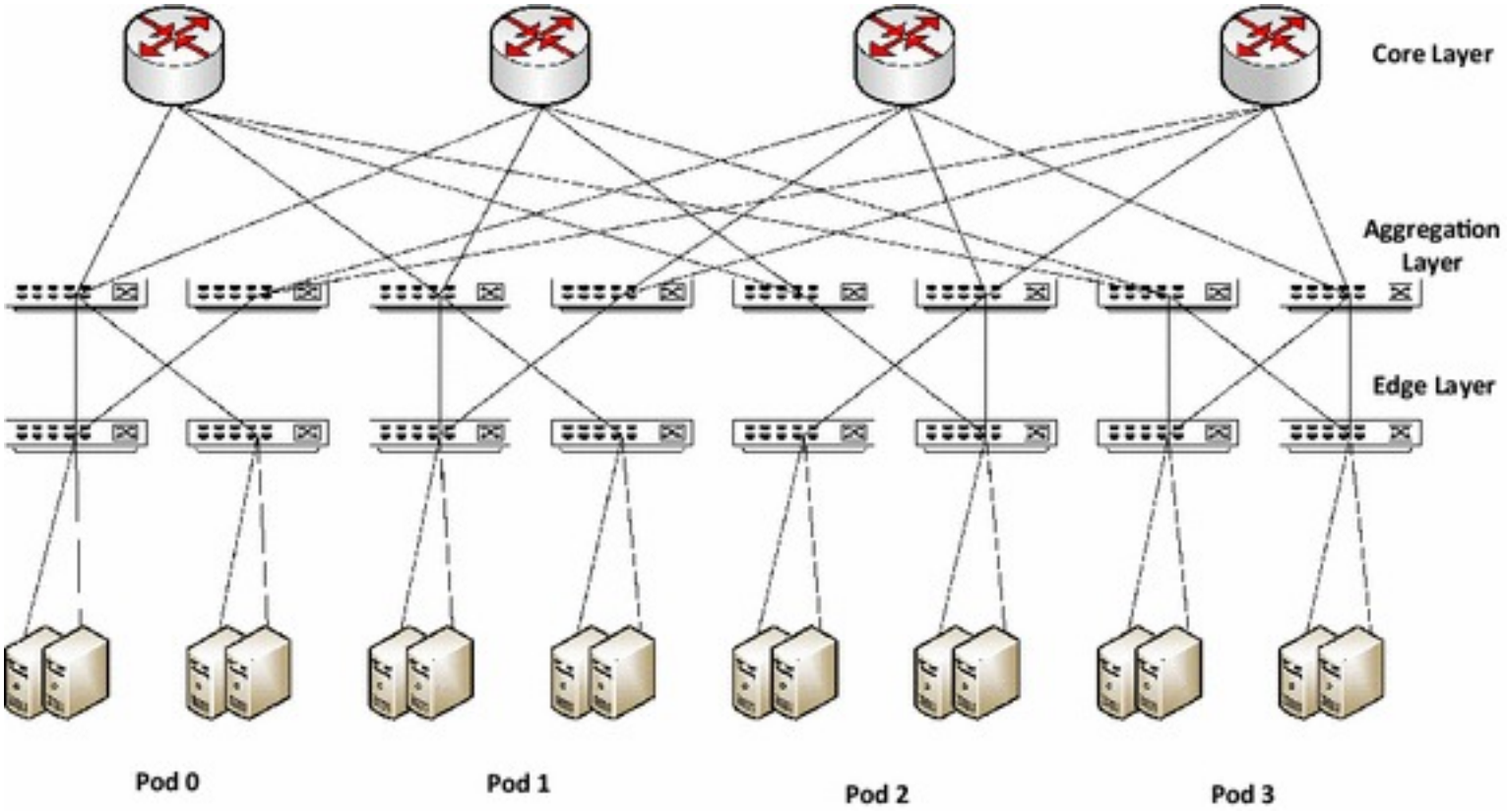
- Topologia 3 tier (menos switches, menos tolerante a falhas)



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4912547/figure/Fig1/

Redes de datacenters

- Topologia Fat-tree (mais switches, mais tolerante a falhas)



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4912547/figure/Fig2/

Estrutura da Internet

- Importante para estimar desempenho das comunicações
  - Atraso médio num caminho entre origem e destino
  - Número de roteadores entre origem e destino (roteador também pode ser chamado de salto ou hop)
- Muita relação com roteamento mas também existem outras “visões” relacionadas com as conexões físicas (cabos marinhos por exemplo) ou com as conexões lógicas (links entre páginas web)

Estrutura da Internet

- No geral as decisões de roteamento para criar um caminho e estabelecer a estrutura são feitas com base nos prefixos de endereço IP pois um prefixo costuma pertencer à mesma organização física
  - ex.: Endereço IP: 192.168.0.1 Máscara: 255.255.255.0. Nesse caso o prefixo é 192.168.0 e esse valor seria usado para definir uma rota. O último número (.1) identifica a máquina. O primeiro número (192.168.0) identifica a rede

Estrutura da Internet

- O comando whois traz informações sobre um dado endereço IP (a quem pertence, qual o prefixo, etc...) conforme registrado na IANA que é quem administra as alocações de endereço IP no mundo
  - O comando costuma já vir instalado em máquinas Unix-like
  - Sintaxe no shell para pegar informações de um IP da USP:
    - whois 143.107.45.22

Estrutura da Internet

- O comando traceroute permite traçar a rota até um dado destino na Internet
  - O comando costuma já vir instalado em máquinas Unix-like
  - Sintaxe no shell para obter o caminho até o servidor web do IME:
    - traceroute www.ime.usp.br
    - Alguns roteadores podem não responder os pacotes do traceroute e por isso o caracter '*' deve começar a aparecer na saída do comando a partir de um determinado hop

Estrutura da Internet

- Mapa lógico da Internet (links entre páginas): https://internet-map.net/
- Mapa físico da Internet (cabos submarinos): https://www.submarinecablemap.com/
- Informação geolocalizada do seu endereço IP: https://whatismyipaddress.com/

Referências

- Raj Jain. The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling. Wiley, 1991.
- Wikipedia, “Detecção de anomalias”, https://pt.wikipedia.org/wiki/Detec%C3%A7%C3%A3o_de_anomalias
- Wikipedia, “Outlier”, https://en.wikipedia.org/wiki/Outlier
- Saed Saya, “Model evaluation”, https://www.saedsayad.com/model_evaluation_c.htm
- Bruno B. Zarpelão, Leonardo S. Mendes, Taufik Abrão, Lucas D. H. Sampaio, Moises F. Lima e Mario L. Proença Jr., “Detecção de Anomalias em Redes de Computadores”, XXVII SBrT, 2009., http://www.uel.br/grupo-pesquisa/secmq/artigos/SBrT-2009.pdf
- Wikipedia, “Data center network architectures”, https://en.wikipedia.org/wiki/Data_center_network_architectures