Usando visualização para documentação rápida de incidentes de segurança

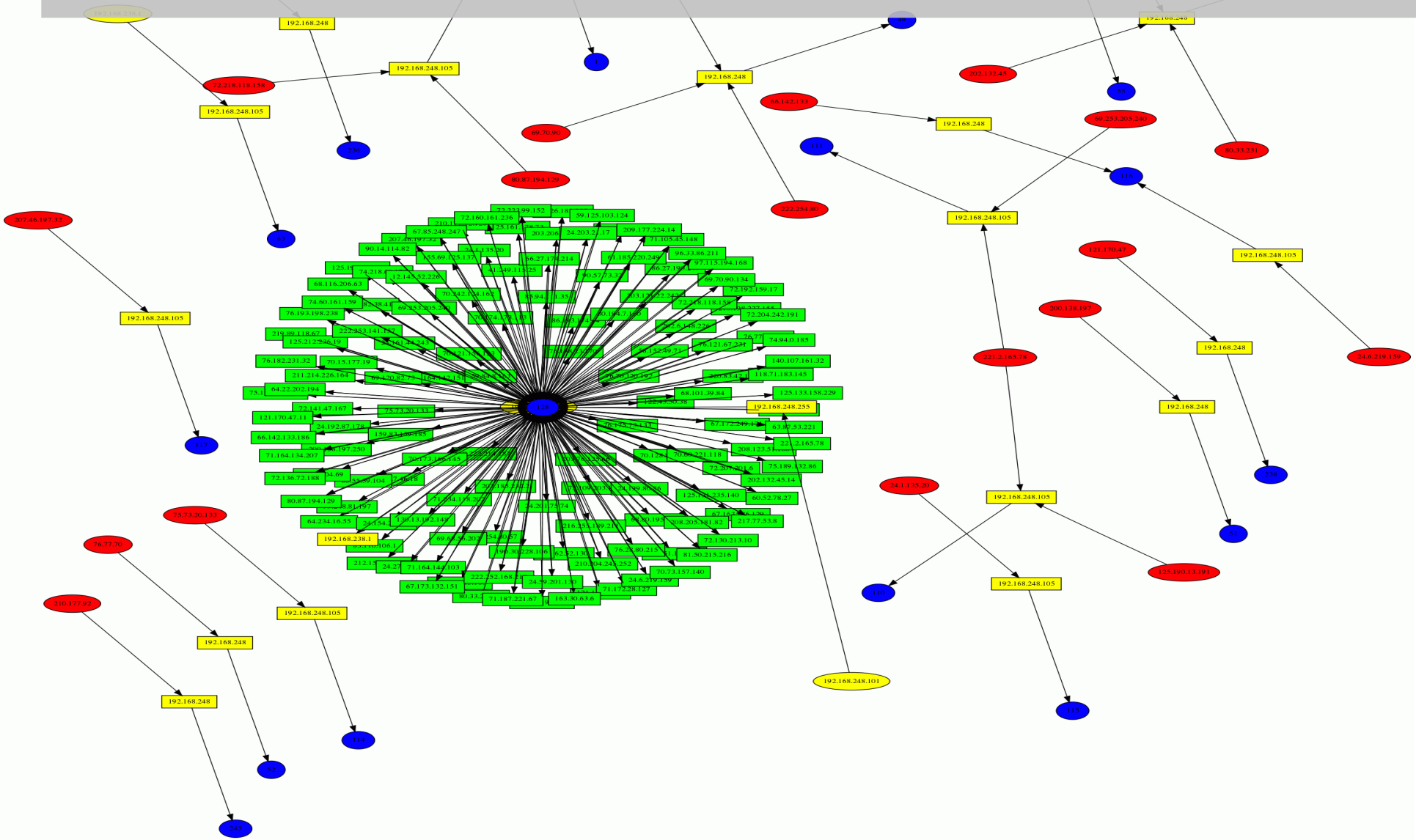
Gabriel Dieterich Cavalcante
Paulo Lício de Geus
LAS – IC – UNICAMP

GTS 16 – São Leopoldo, 2010

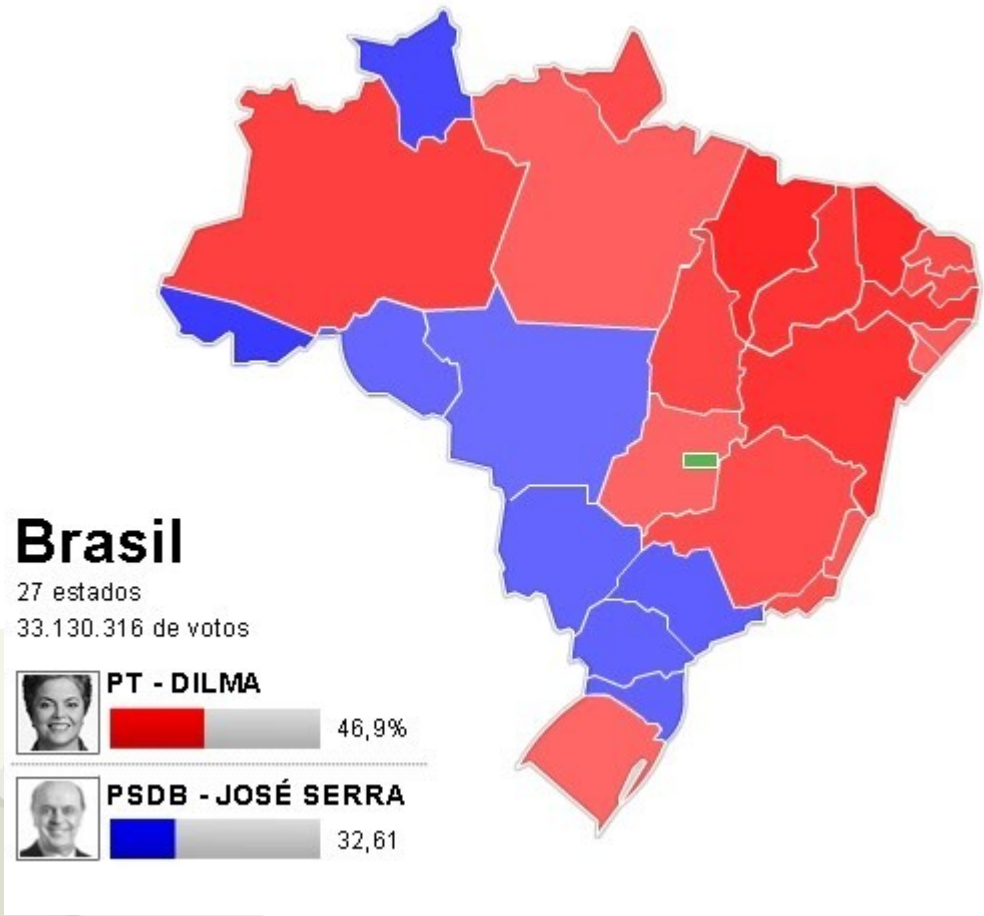
- Moro em Campinas – SP
- Sou Doutorando em Ciência da Computação pelo IC – UNICAMP
- Mestre pelo mesmo Instituto.
- Usuário Linux desde 2003/2004 (feliz com gentoo desde 2007)
- Trabalho no LAS – IC -UNICAMP
  - Verificação de Integridade;
  - Forense;
  - Sanitização de dados (anonimização de pacotes);
  - Security Visualization;
  - 2 vezes aluno do Google Summer of Code;
- Mantenedor do Picviz, desenvolvedor loganon.

Introdução

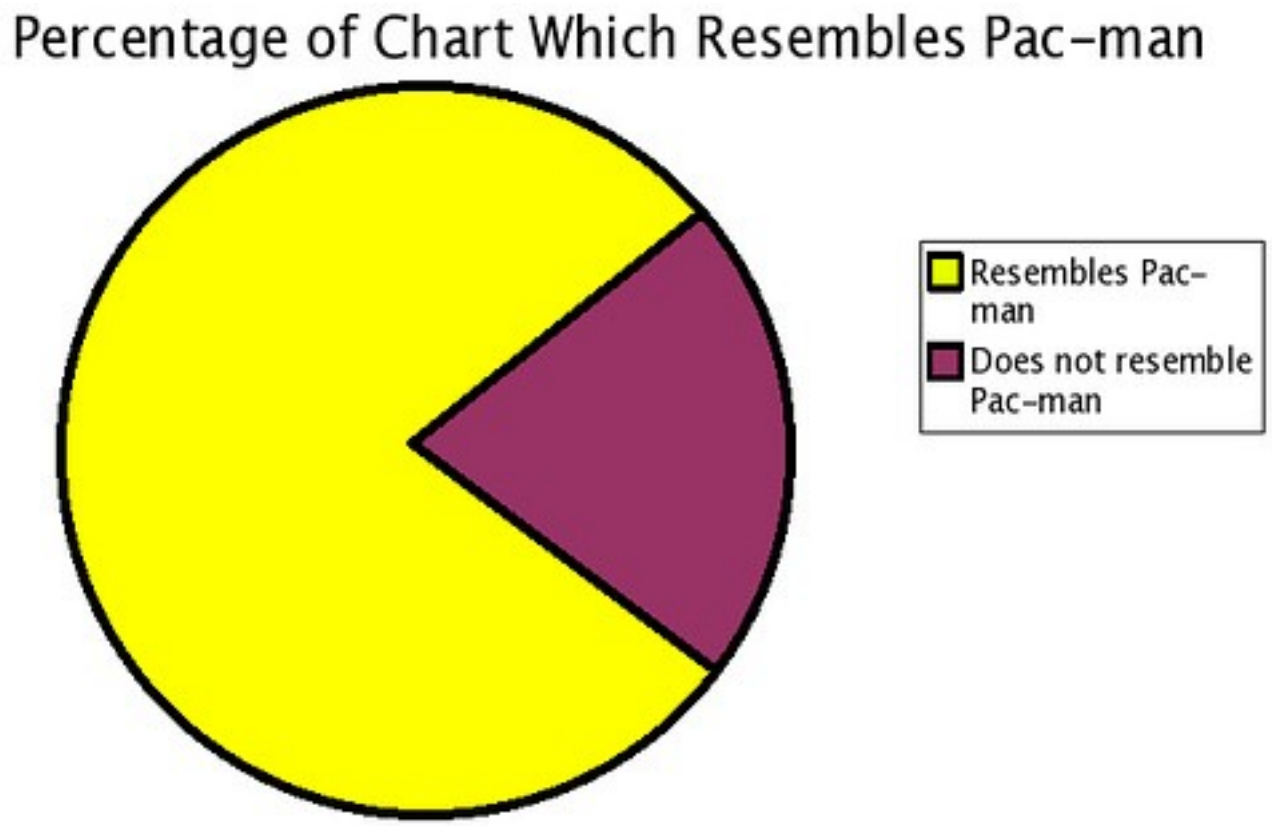
- Visualização? Pra quê?
  - Explorar e descobrir...
    - ➔ Às vezes para responder perguntas (ou trazer novas)



- Ajuda nas decisões, comunicar informações..
- Melhorar a eficiência (da análise, do conteúdo analisado).

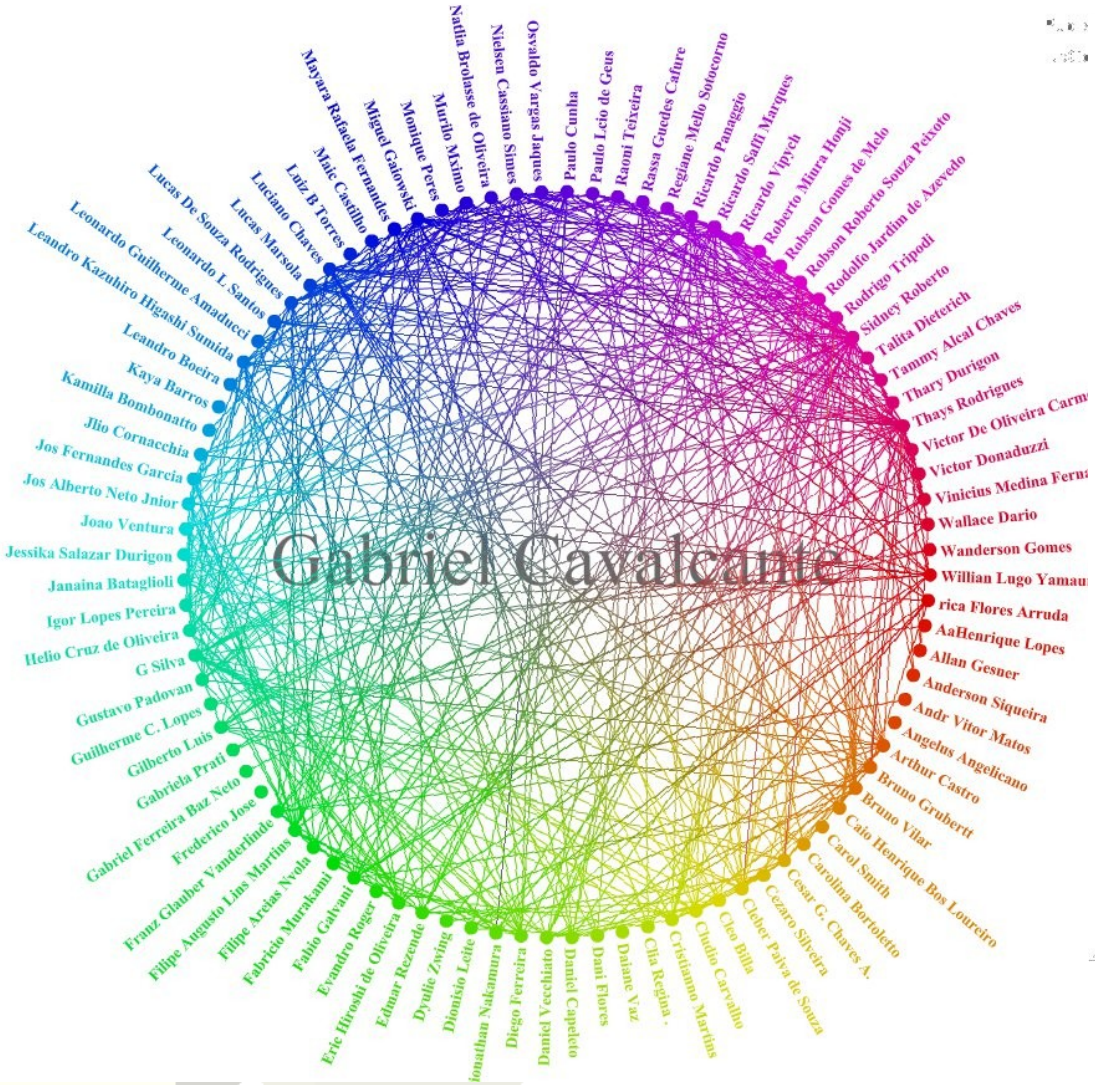


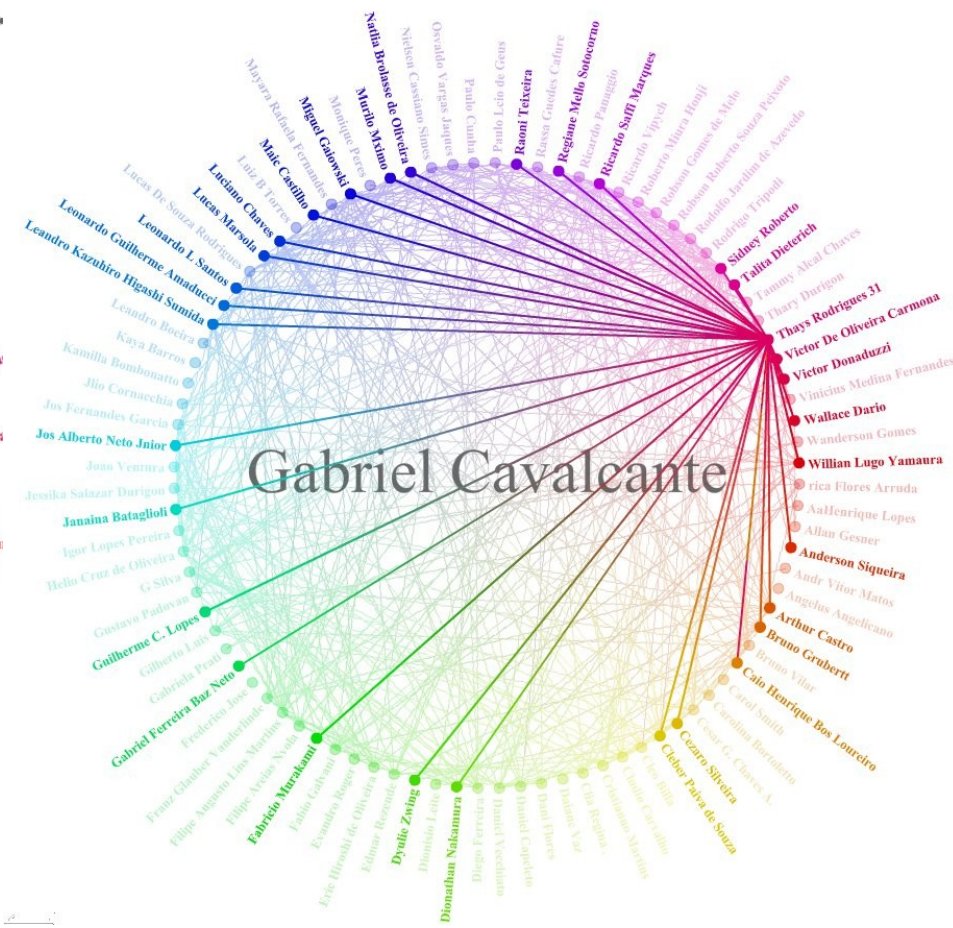
- Diversão


Percentage of Chart Which Resembles Pac-man
Resembles Pac-man
Does not resemble Pac-man

Introdução

- Às vezes por pura necessidade...



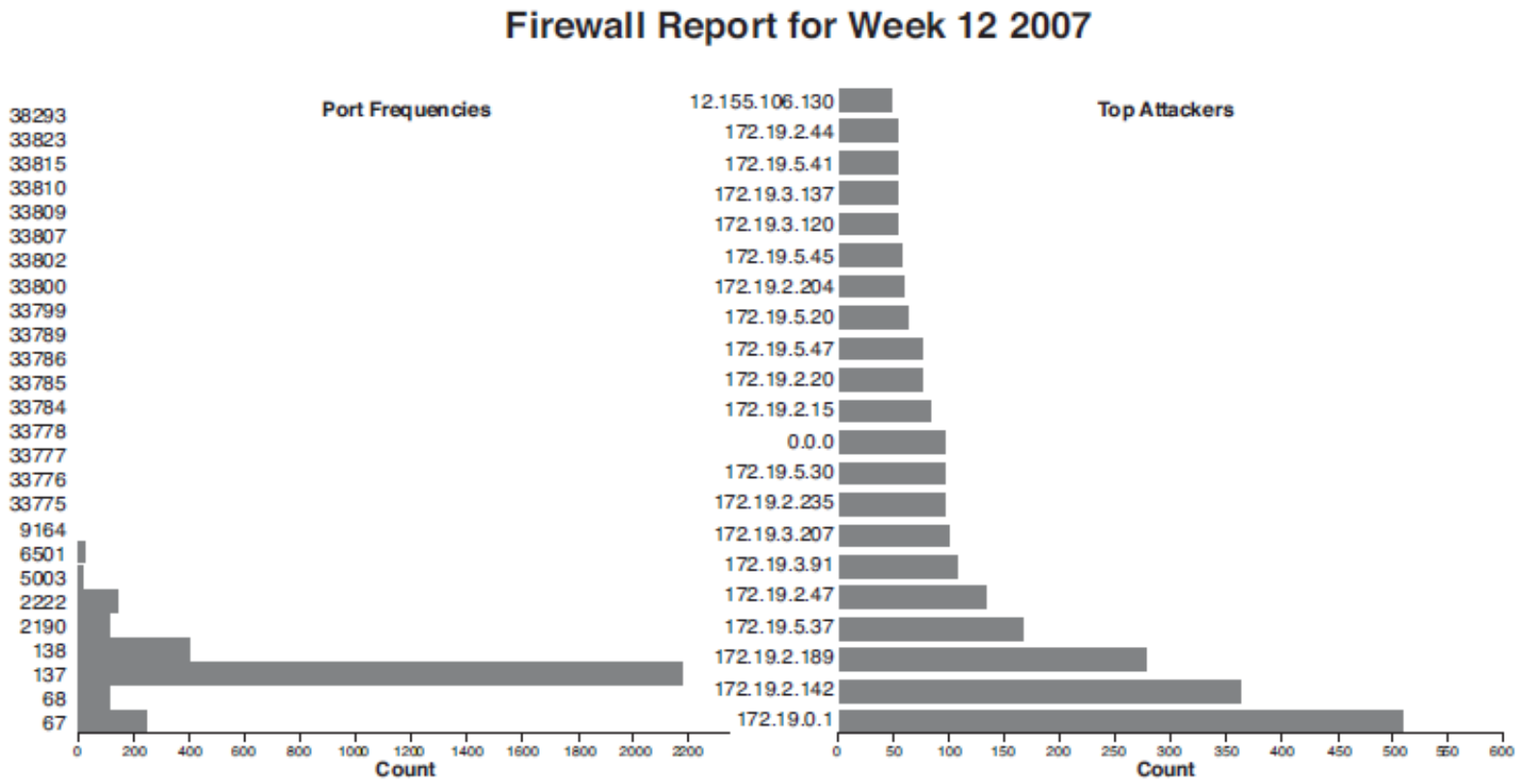


- Área relativamente nova:
  - A quantidade de dados que precisam ser analizados em problemas de segurança..
  - Quantidade de dados armazenados (*data-collected*) em ambientes de TI é cada vez maior;
  - Deptos. Regulatórios fazem pedidos de análises regulares (alguns países);
  - Foco de ataque está acima da camada de rede (*network stack*).
    - → Aplicações geram muitos eventos (normalmente conservadoras).

- Classes de Visualização
  - Relatórios:
    - Gráficos simples;
    - Usados para melhorar a qualidade técnica do relatório;
    - **Alvo**: em grande maioria **leigos****.
  - Análise Histórica (*after attack*):
    - *Time-series analysis*;
    - Correlação;
    - Gráficos interativos (análise);
    - Análise forense;
    - **Alvo**: Analistas e *report-makers*, em suma *somos nozes* ;-)
  - Monitoramento e Análise em Tempo real:
    - Dashboards:
      - Operacional
      - Tático
      - Estratégico
    - **Alvo**: Times de monitoramento (*datacenters)*;

Security Visualization - Reporting

- Melhor ferramenta para comunicar e resumir dados:
  - Ex: Tráfego bloqueado durante os últimos sete dias.



- Normalmente os dados são processados manualmente;
- Gráficos simples: deve-se entender exatemente o que se passa.

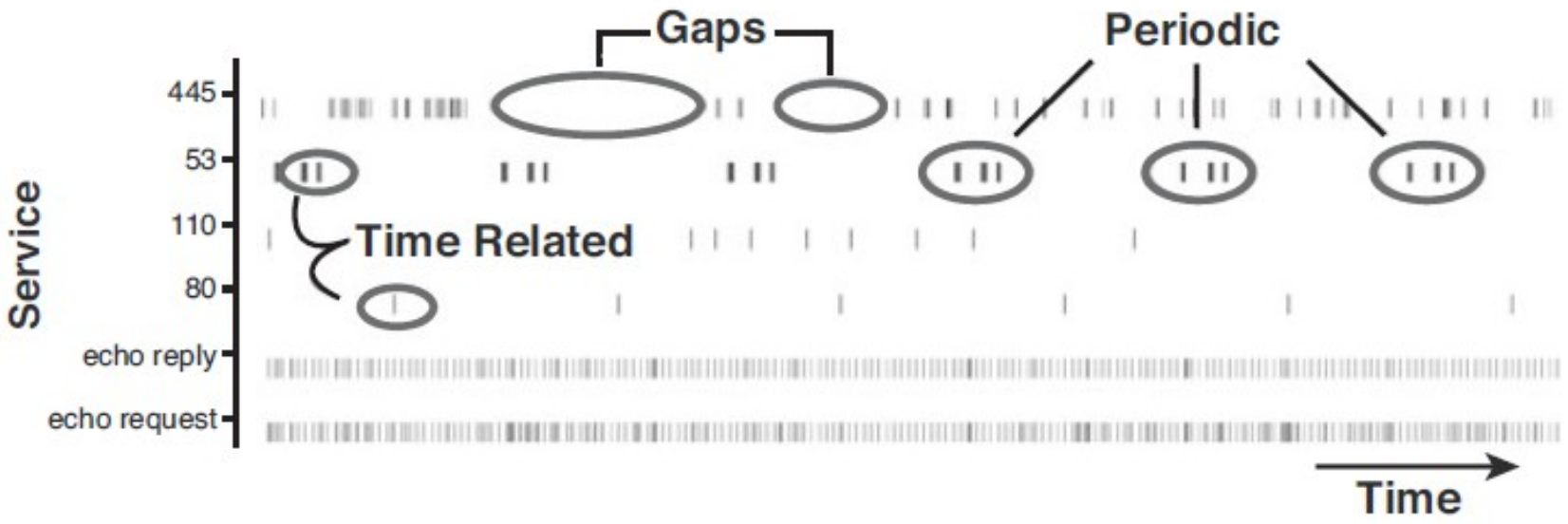
Reporting tools

- Propósito Geral:
  - OpenOffice, Bill's Office..
  - CrystalReport;
  - GnuPlot;
  - Problema: Como representar campos mais complexos, IPS p. ex.?

- Security Reporting Solutions:
  - Geralmente pagas (e nada baratas);
  - Centralizam logs, fazem toda a manipulação e preparam relatórios baseados em *templates*.

- Bibliotecas de programação:
  - Java (eca!), Php, perl, python, C entre outros..
  - *Frameworks* de visualização;

- *Time-series* (dados coletados ao longo de um período):
  Ex. Todos os logins em uma máquina (hora+login).
  Objetivo:
  - Predizer valores, analisar as variâncias de eventos durante o período, anomalias, tendências..

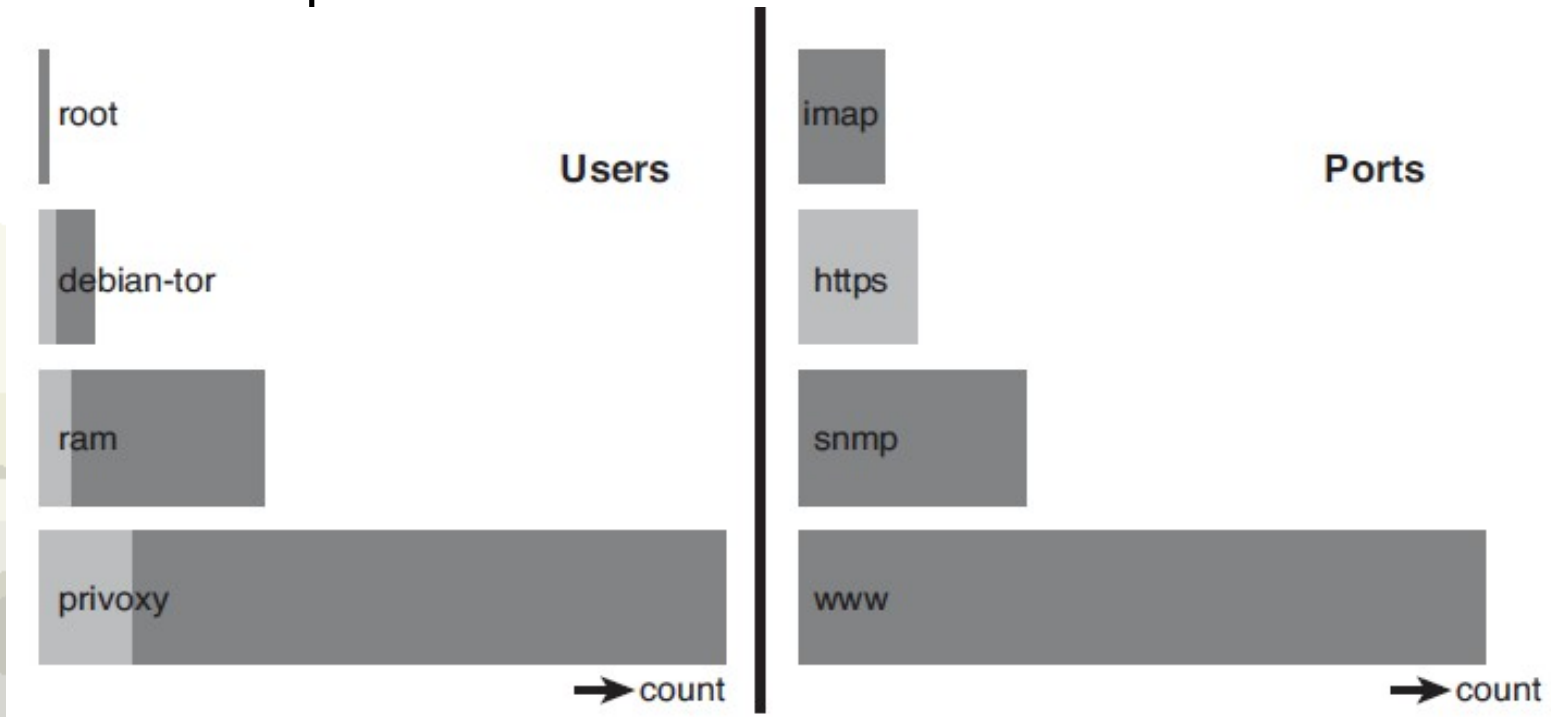


- Interactive Analysis
  - Até agora vemos somente imagens ou gráficos estáticos..
  - Normalmente durante a preparação definimos suas propriedades:
    - Cor, tamanho, tipo, transparência...
  - Não sabemos se a saída será a melhor...

  - Ferramentas com gráficos interativos (dinâmicos) permitem a análise dos dados plotados:
    - Procurar pelo “Mantra” dos dados:
      1 – *Overview*;
      2 – Mudança nos atributos;
      3 – *Zoom* e filtros;
      4 – Detalhes sob demanda;

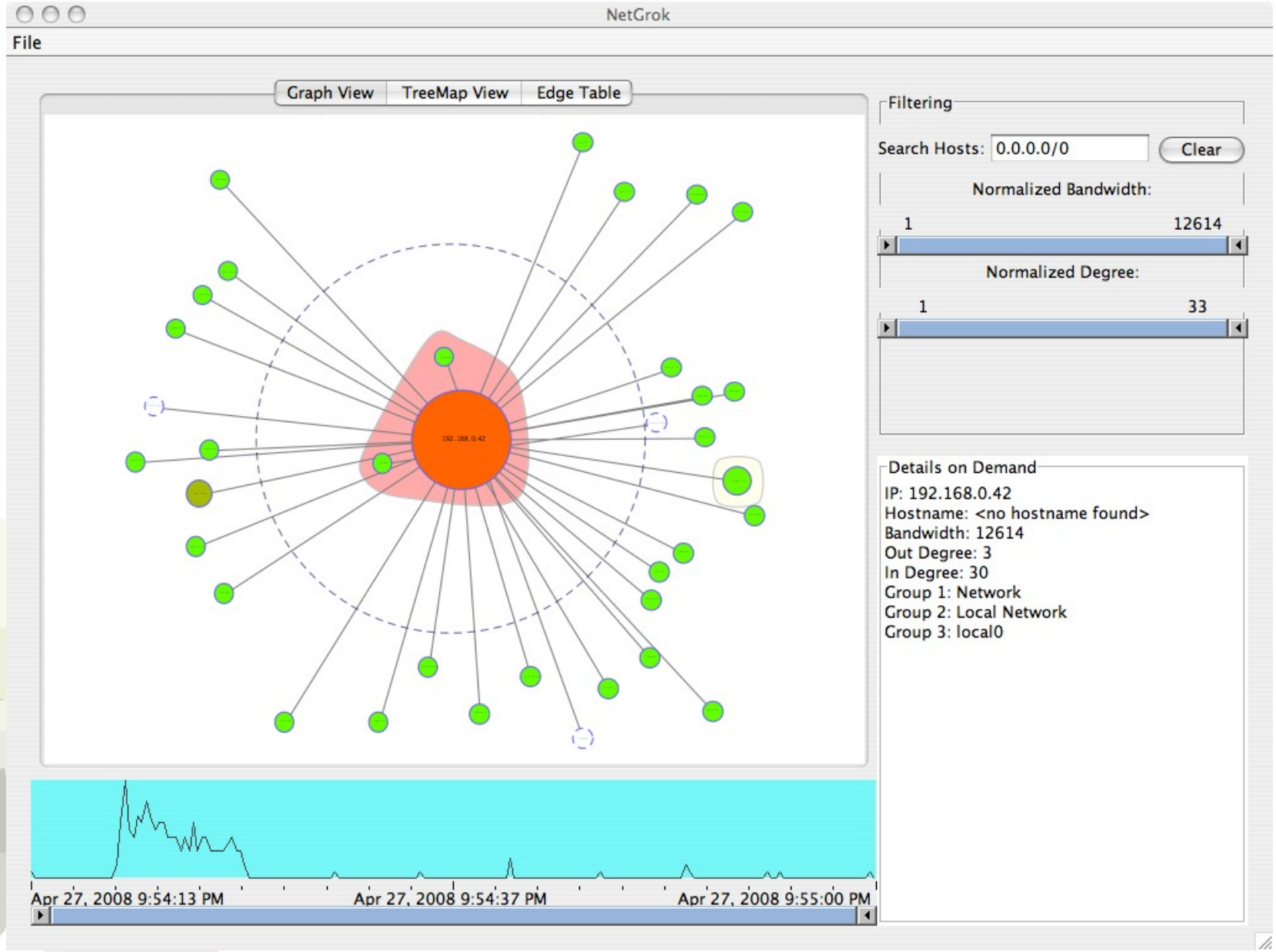
- Interactive Analysis
  - Até agora vemos somente imagens ou gráficos estáticos..
  - Normalmente durante a preparação definimos suas propriedades:
    - → Cor, tamanho, tipo, transparência...
  - Não sabemos se a saída será a melhor...

  - Ferramentas com gráficos interativos (dinâmicos) permitem a análise dos dados plotados:



Interactive Analysis

- NetGrok

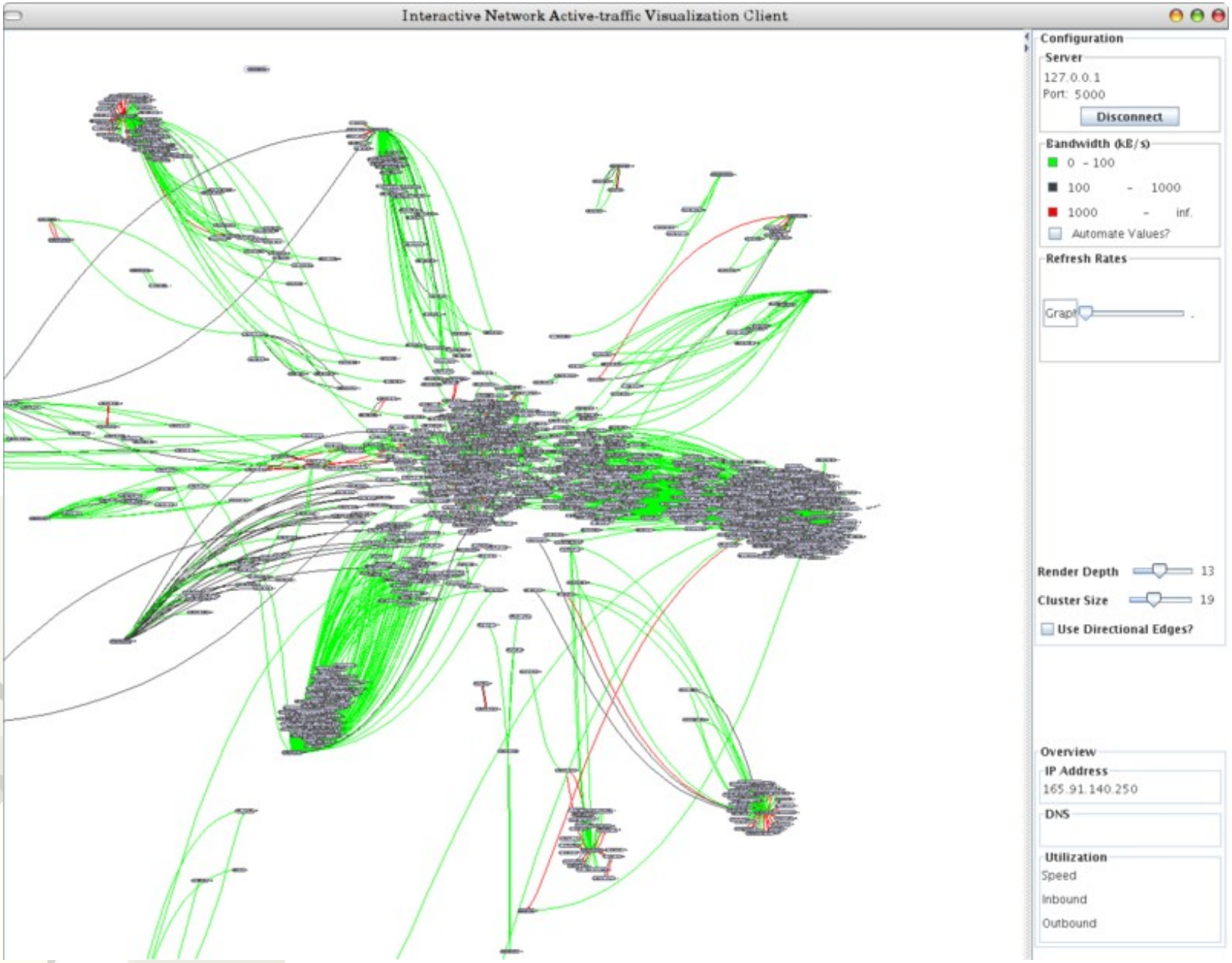


- Logstalgia

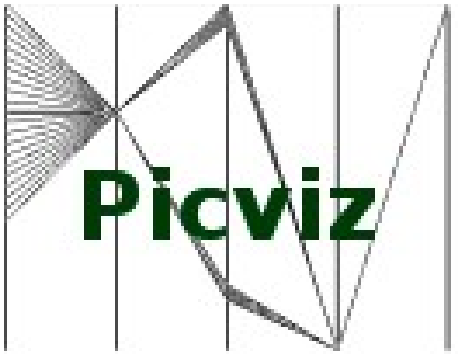

Logstalgia
Website Access Log Visualization
logstalgia.googlecode.com/
200
/js/mer
/js/add
Images
/images
/images
/images
/images
/images
/images
/images
/images
Misc

Interactive Analysis

- INAV



- E muitos outros..
  - ggobi
  - AfterGlow
  - Treemap
  - FlowMatrix
  - Gephi
  - glTail
  - Gource
  - GraphMovie
  - prefuse
  - SilentRunner
  - Maltego
  - ...



“Finding a needle in a haystack...
when you don’t even know how the needle looks like”

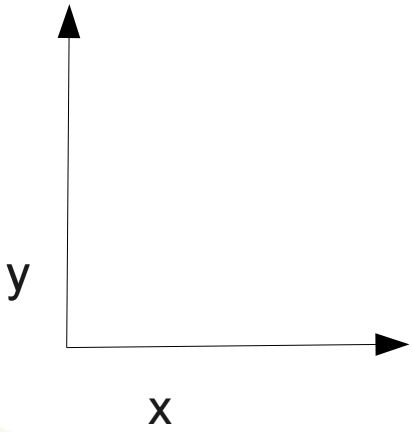
- Cria gráficos a partir de seus *logs* (coordenadas paralelas);
- Plota qualquer dado que possa ser “parseado”;
- Pode lidar com um grandíssimo número de eventos;

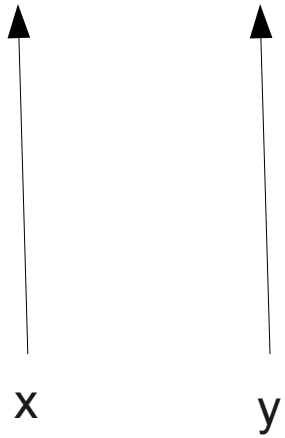
- Êpa, mas que tipo de *Logs*? Qualquer tipo!
  - ***Syslogs***
    - *Nov 6 13:12:04 quine avahi-daemon[2285]: Interface eth0.IPv4 no longer relevant for mDNS.*
    - *Nov 6 13:12:06 quine ifplugd(eth0)[1811]: Program executed successfully.*
    - *Nov 6 13:12:06 quine kernel: ADDRCONF(NETDEV_UP): eth0: link is not ready*
    - *Nov 6 13:12:24 quine kernel: Unhandled event received : 0x50*
  - ***Database Logs***
    - *sql> SELECT * FROM logdb WHERE user = "ptc";*
  - ***Network***
    - *08:50:01.522077 arp who-has 10.0.0.254 tell 10.0.0.1*
    - *08:50:01.522115 arp reply 10.0.0.254 is-at 00:69:de:ad:be:ef*
    - *08:50:01.522210 IP 192.168.0.1.5860 > 172.16.17.235.33373: UDP, length 25*
    - *08:50:01.522377 IP 192.168.0.1.5860 >10.30.254.247.18946: UDP, length 25*
  - ***Outros***
    - Análise estática de *malwares*, comandos executados em uma máquina,
    - etc.

- Cenário
  - Muita, muita mas muuuita informação (dados);
  - A estrutura dos logs depende do serviço/aplicação;
  - Interação dificultada;
  - Quando automatizada se baseia em “greps” (pcre based);
  - 1 única máquina consegue produzir milhares de linhas.

- *Overdose* de *logs*
  - A maioria das pessoas está satisfeita apenas em olhar para eles;
  - As empresas/organizações estão cheia de pessoas ocupadas (preguiçosas?);
  - Sem brincadeiras, analisar assim é escravidão;

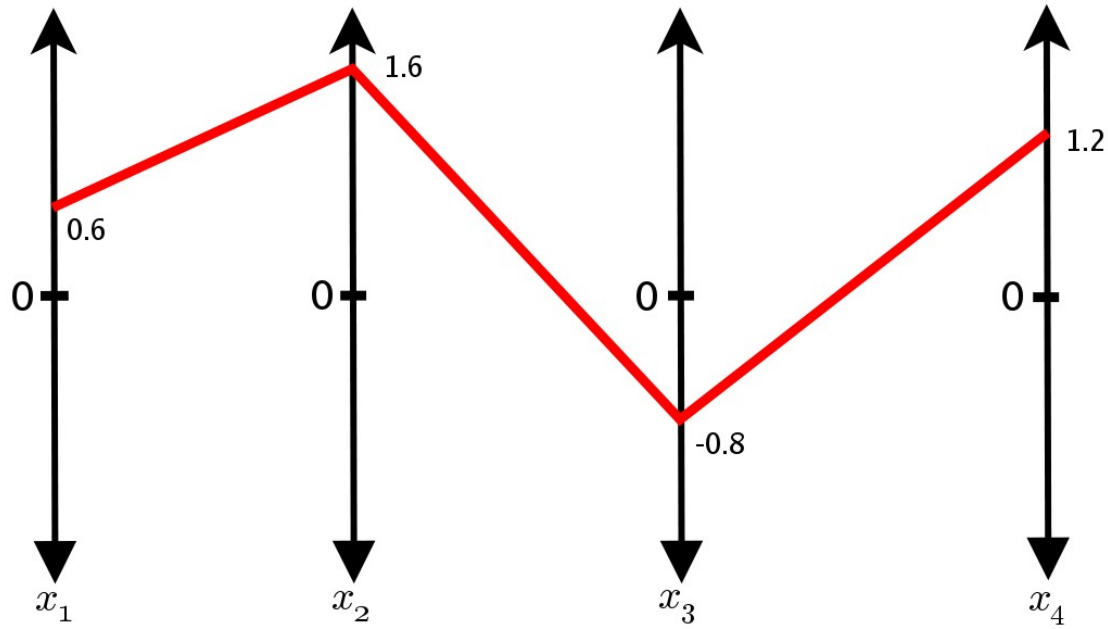
- Uso de coordenadas paralelas:



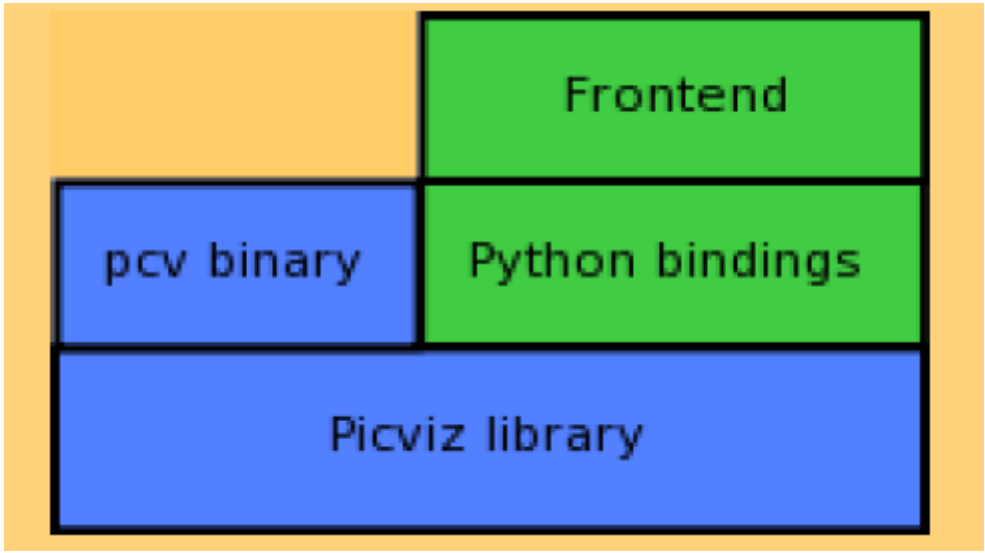


- Propriedades:
  - N-dimensões: um eixo por dimensão;
  - Eixos equidistantes;
  - Eventos infinitos: uma linha por evento;

$\vec{v} = (0.6, 1.6, -0.8, 1.2) \in \mathbb{R}^4$



- Arquitetura:
  - *Scripts*: Transformação de logs em *PicViz Graph Description Language* (PGDL)
  - pcv: Linha de comando para transformar um PGDL em imagem estática
  - picviz-gui: *Frontend* para manipulação do gráfico



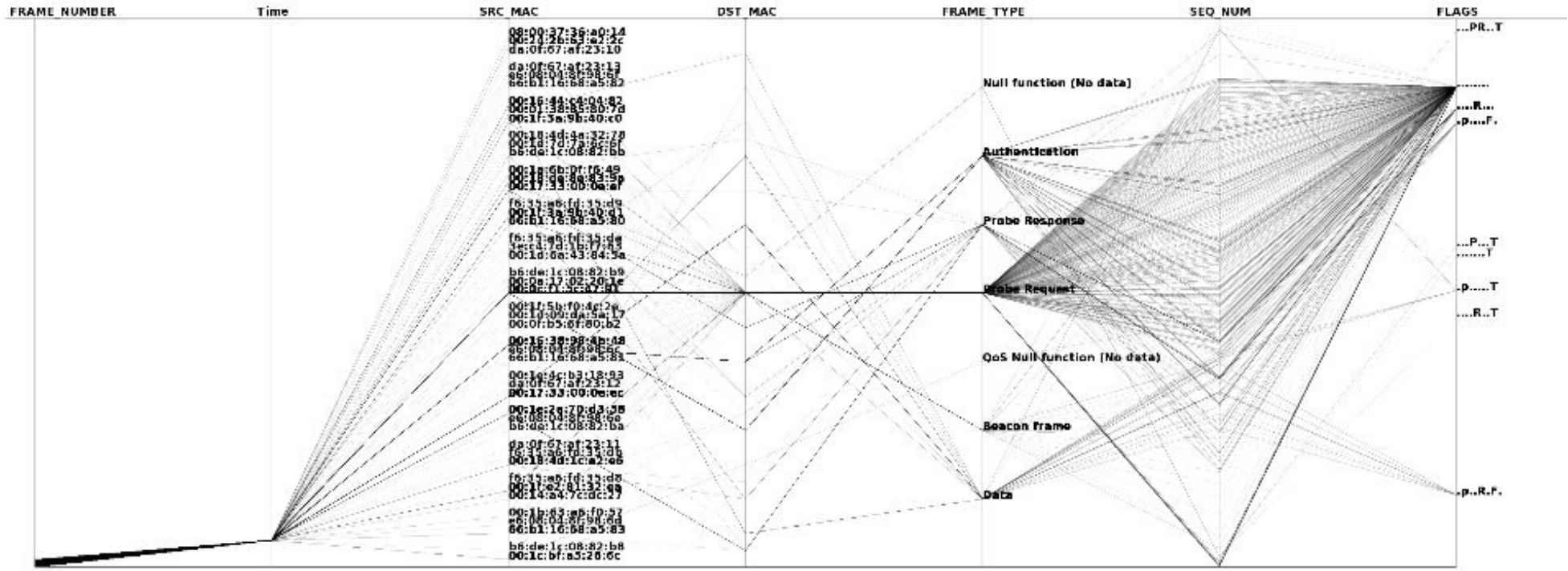
Parsing

Nov 26 17:27:16 [sshd] Accepted keyboard-interactive/pam for fernando from 200.230.121.X port 56162
Nov 26 17:28:49 [sshd] Received disconnect from 10.2.1.243: 11: disconnected by user gabriel
Nov 26 17:28:58 [sshd] Failed keyboard−interactive/pam for invalid user paul from 200.216.236.X port 50108

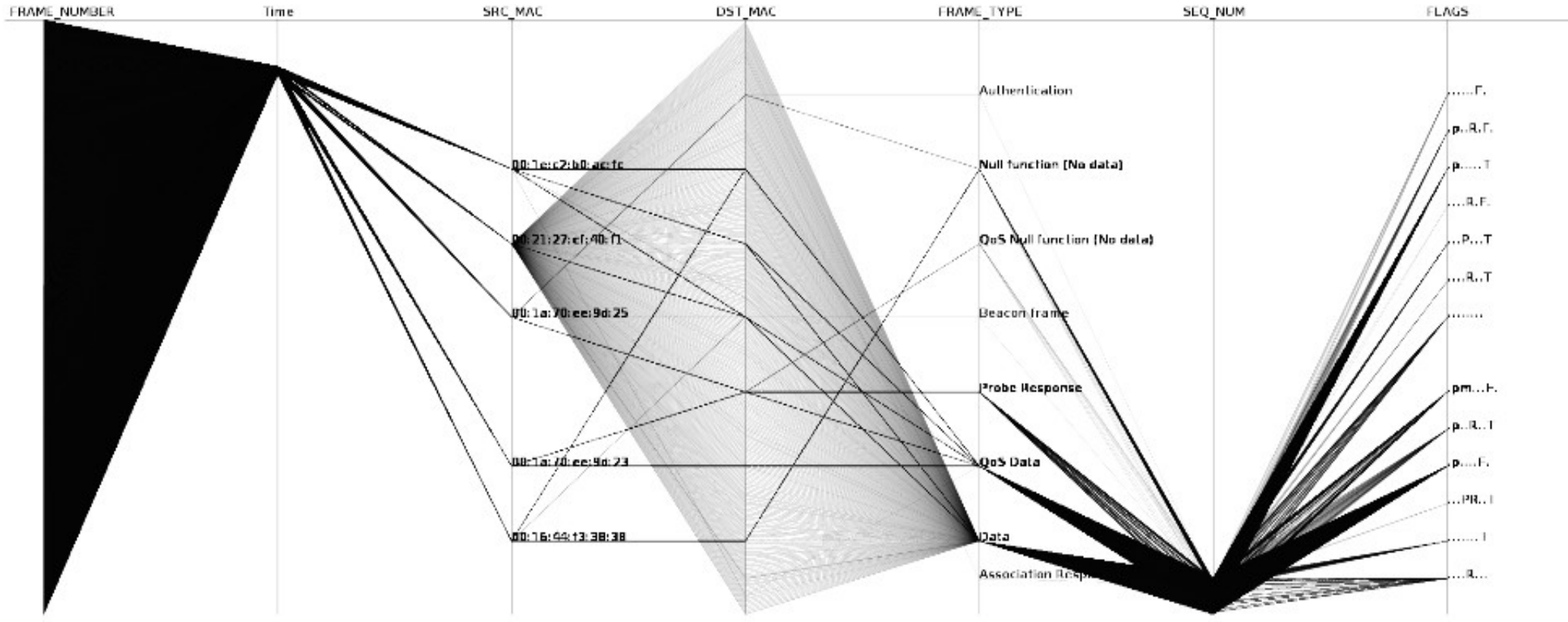
PGDL

```
header {
    title = "SSH Logins";
}
axes{
    enum day[label=”Dias”];
    timeline time;
    string auth;
    ipv4 src;
    enum login;
}
data{
    day="26" , time="17:27:16" , auth="Accepted keyboard-interactive/pam" ,
      src="200.230.121.X", login="fernando";
    days="26" , time="17:28:58" , auth="Invalid user" , src="200.216.236.X",
      login="paul" [penwidth="0.05",color="#FF0000"];
}
```

- O Poder das coordenadas paralelas:
  - Dados de uma rede sem-fio usando WEP.



- O Poder das coordenadas paralelas:
  - Dados de uma rede sem-fio sofrendo WEP cracking.

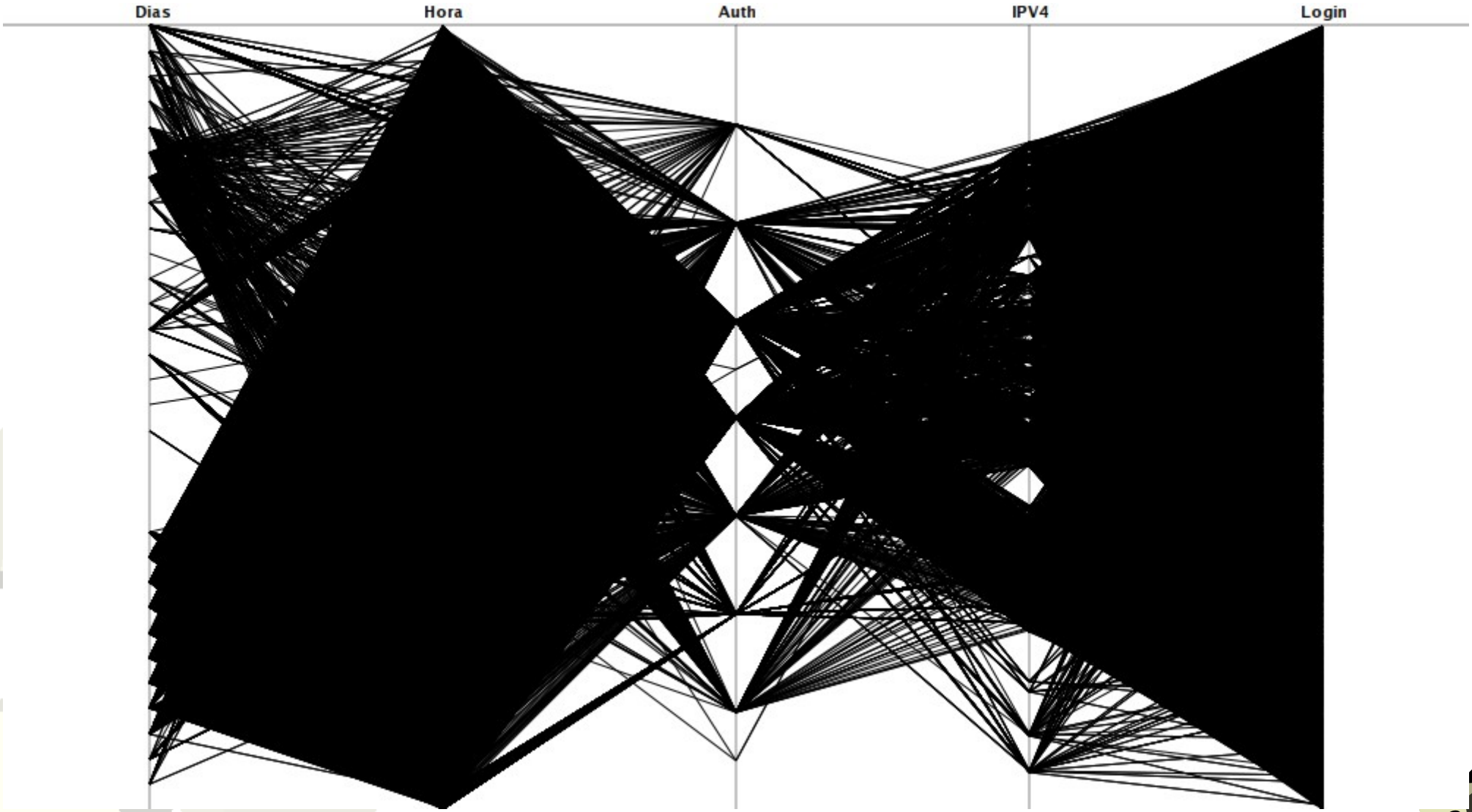


PicViz – Análise do *log* de SSH

- Dados Junho/2010:
  - *sshd* um servidor em algum lugar da USP (sem *denyhosts*);
  - Constatação da ação nada convencional de *bots*;
  - Interação dos dados com a interface gráfica;
  - Criando *layers*, colorindo, mudando a espessura das linhas..

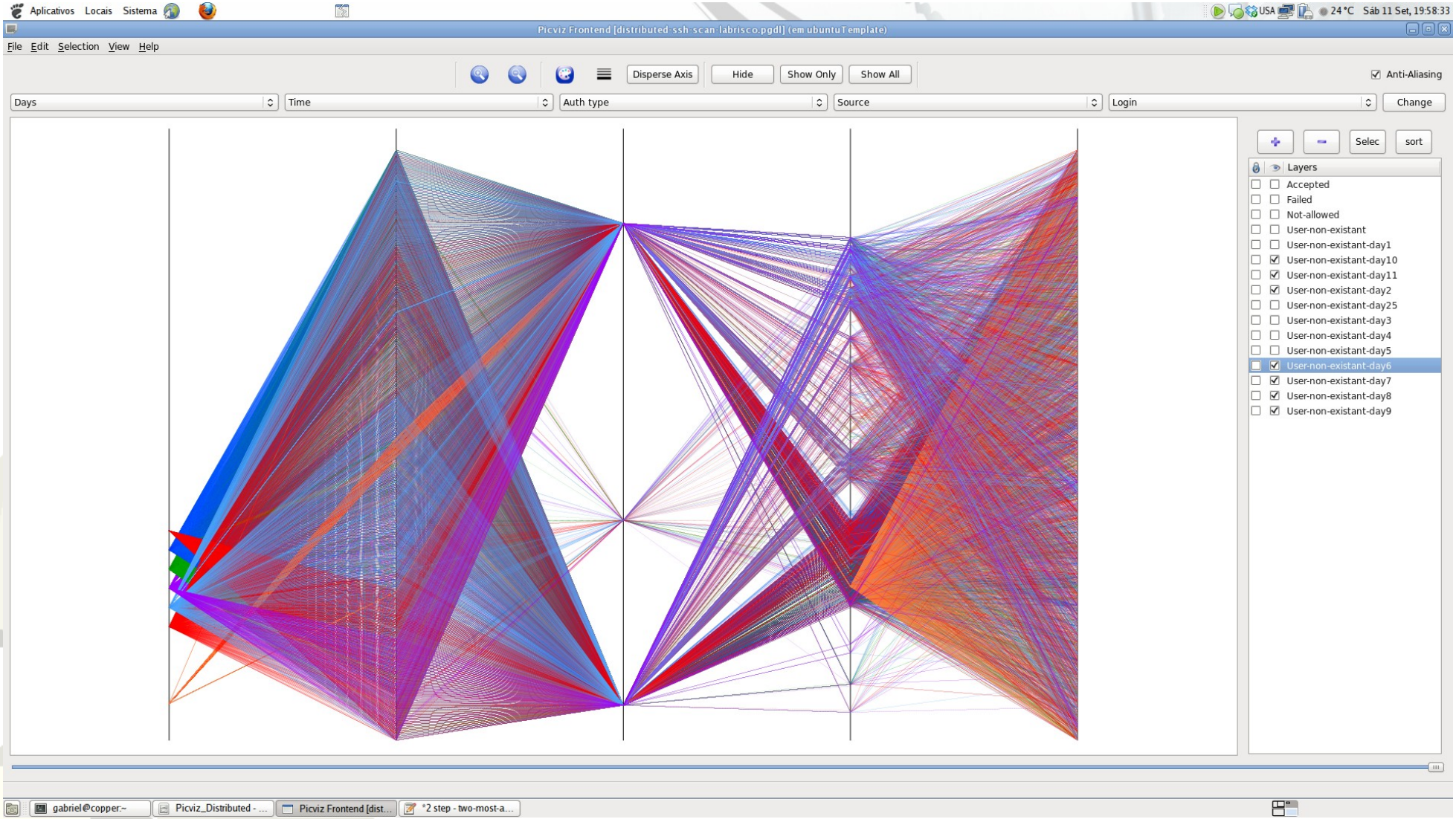
Análise sshd - PicViz

- Imagem esclarecedora:



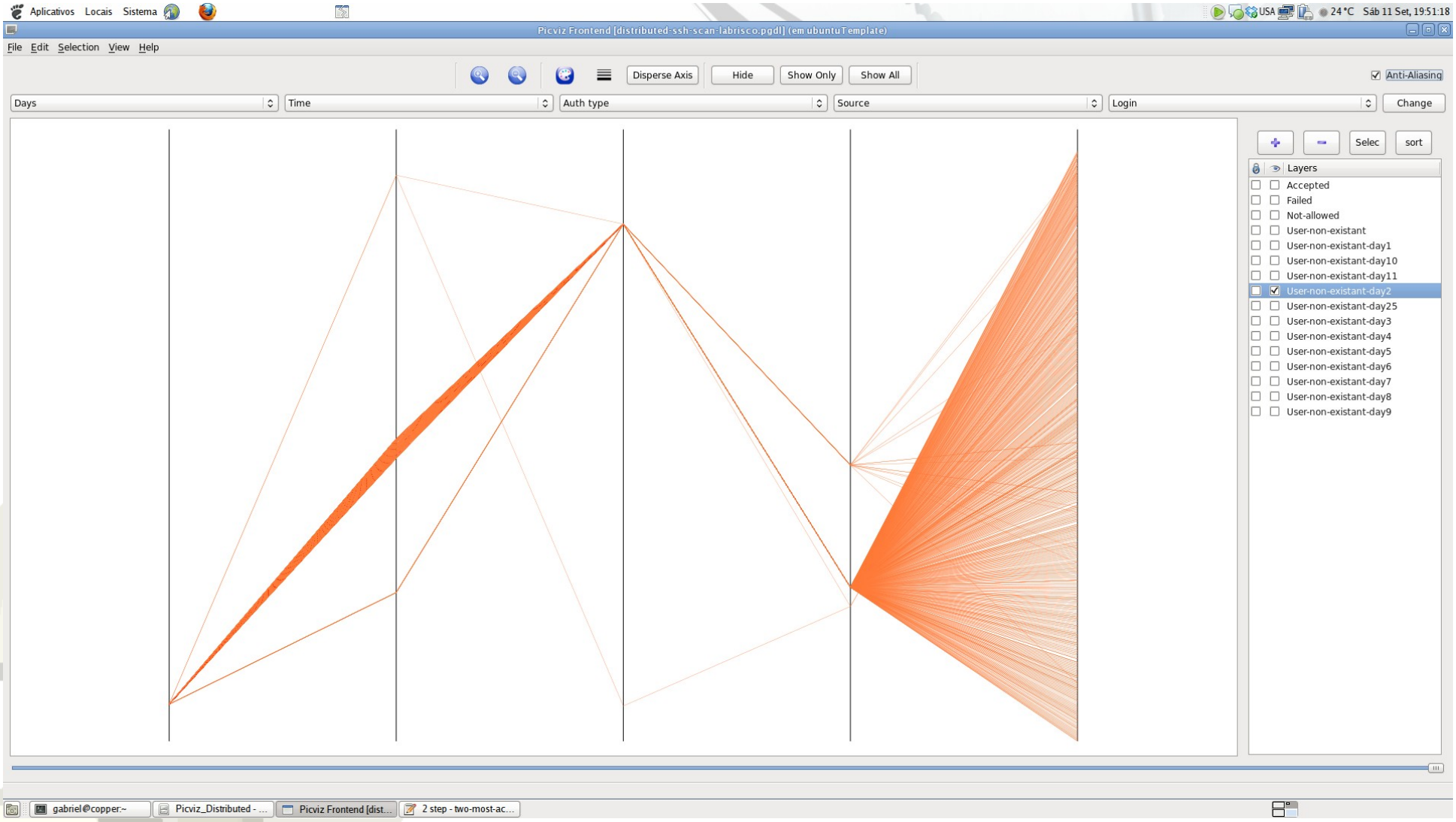
Análise sshd - PicViz

- Isolando os dias mais ativos



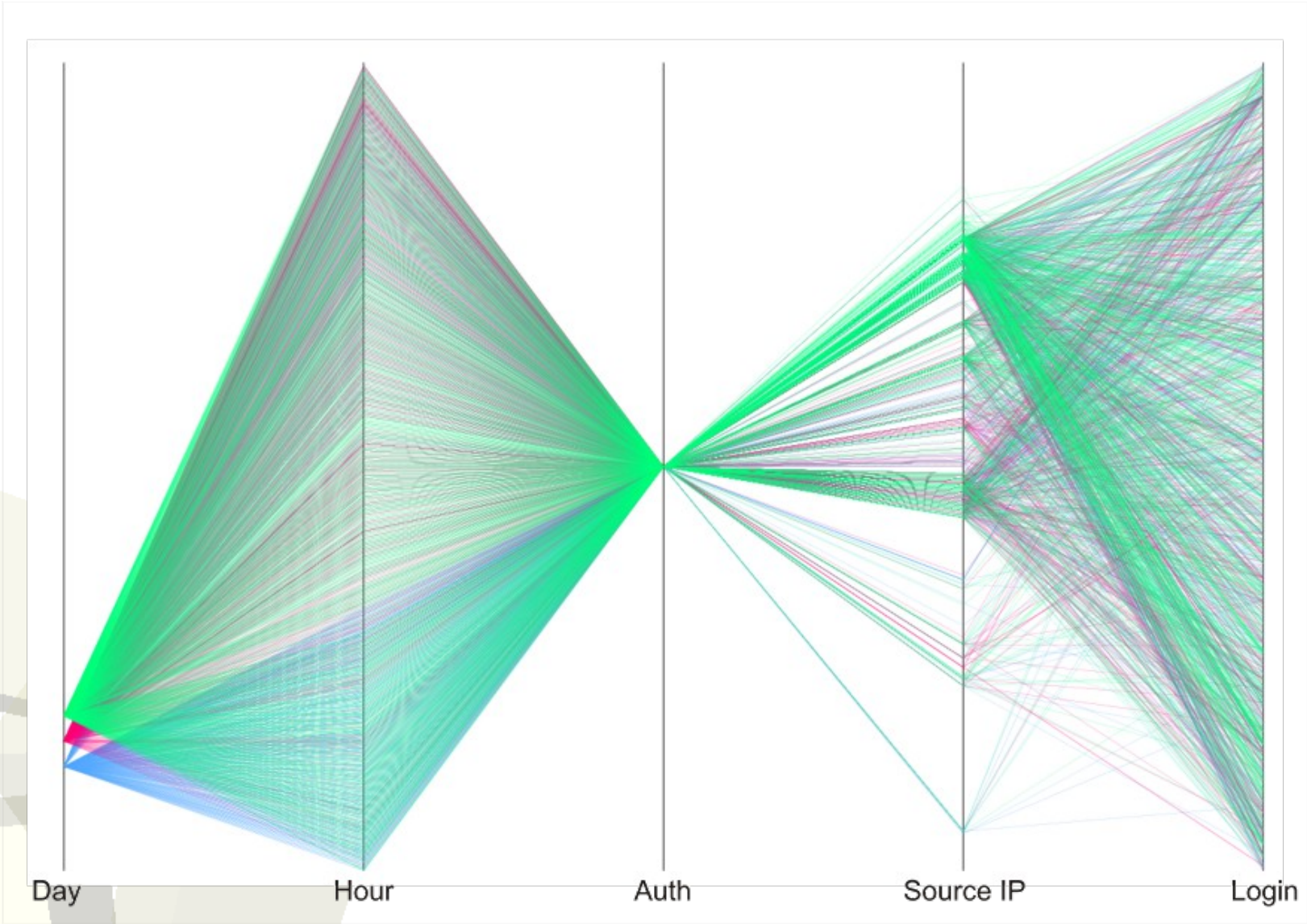
Análise sshd - PicViz

- Dia 2 isolado.



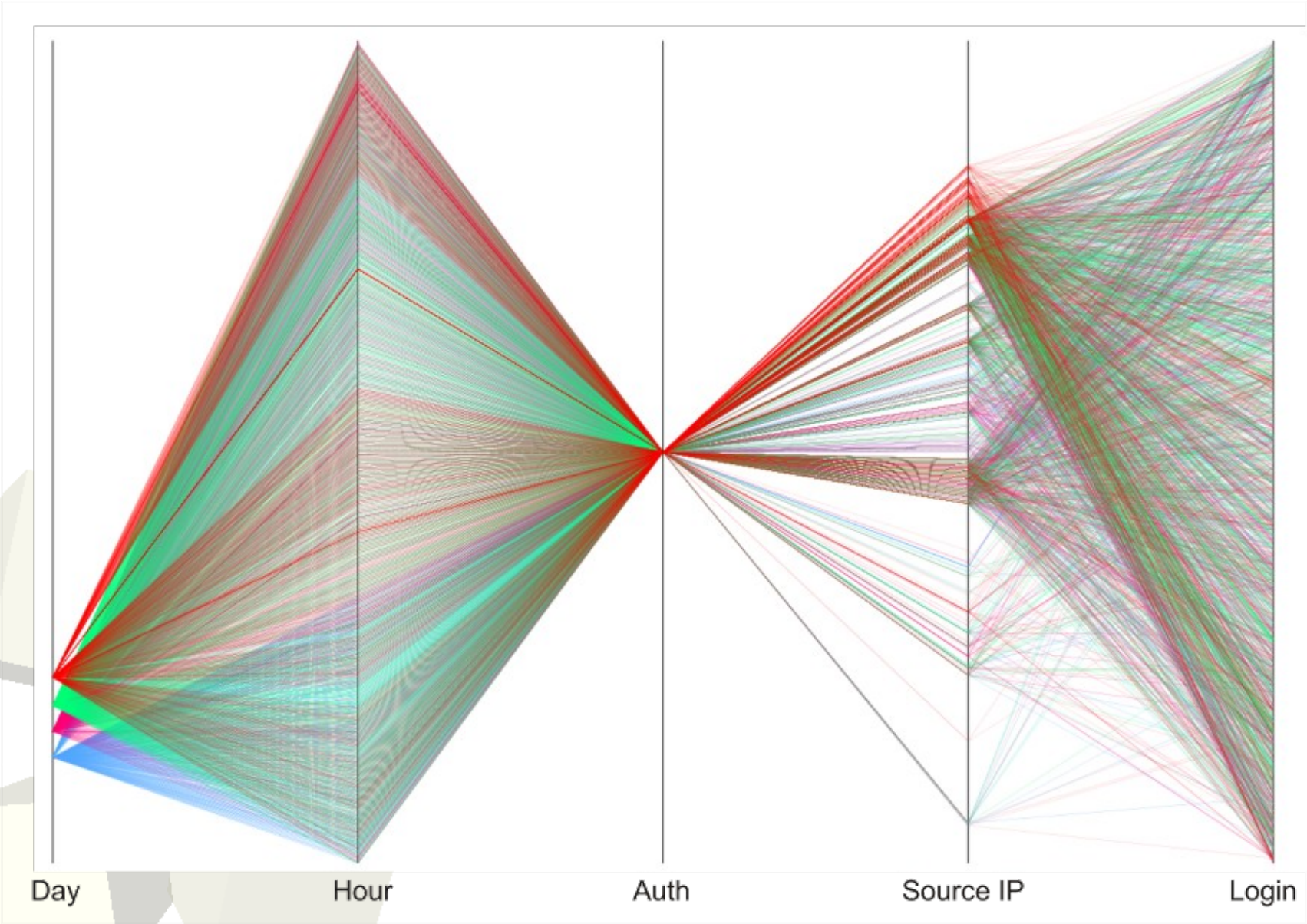
Análise sshd - PicViz

- Dias: 4, 5 e 6.



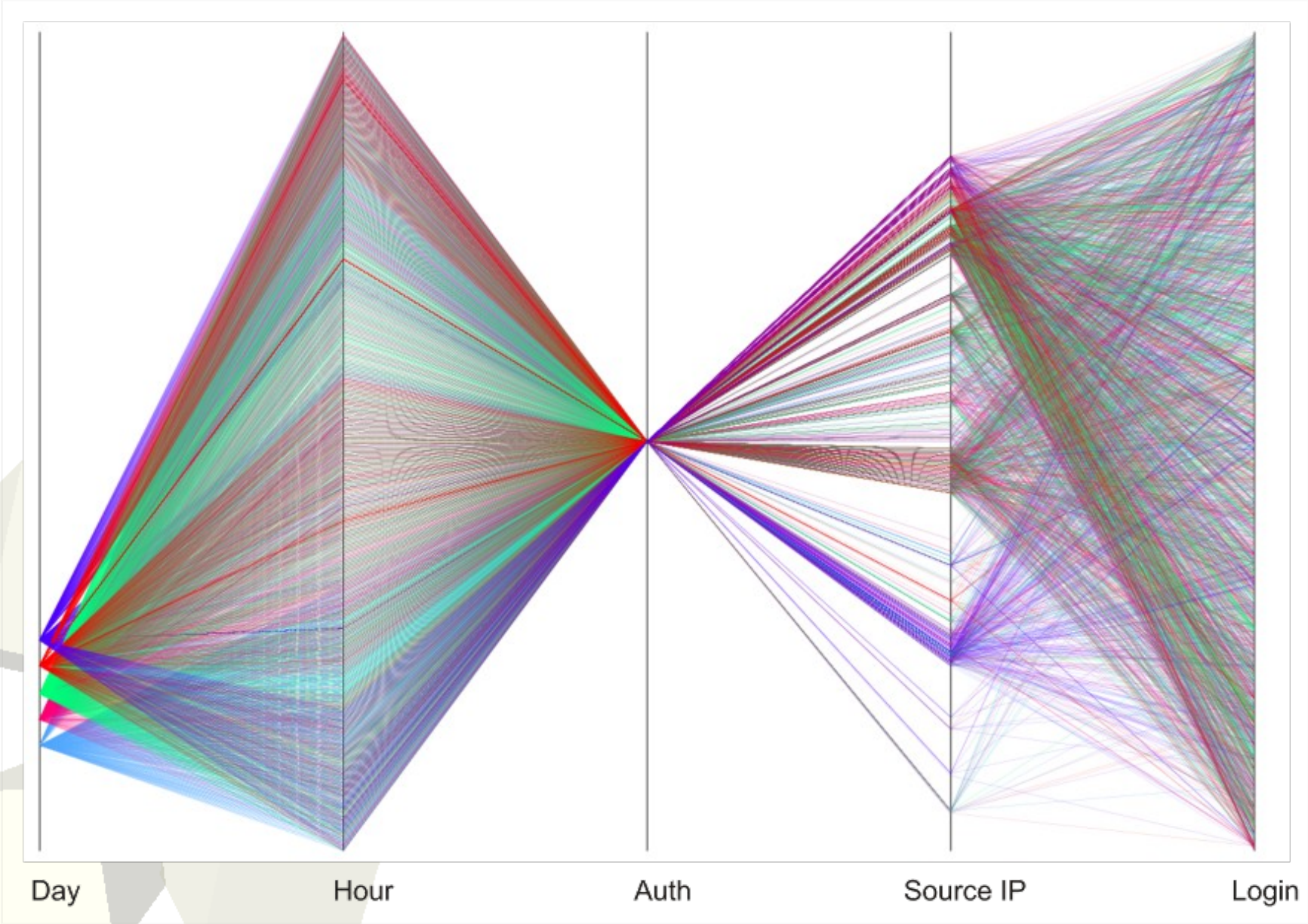
Análise sshd - PicViz

- Dias: 4, 5, 6 e 7.



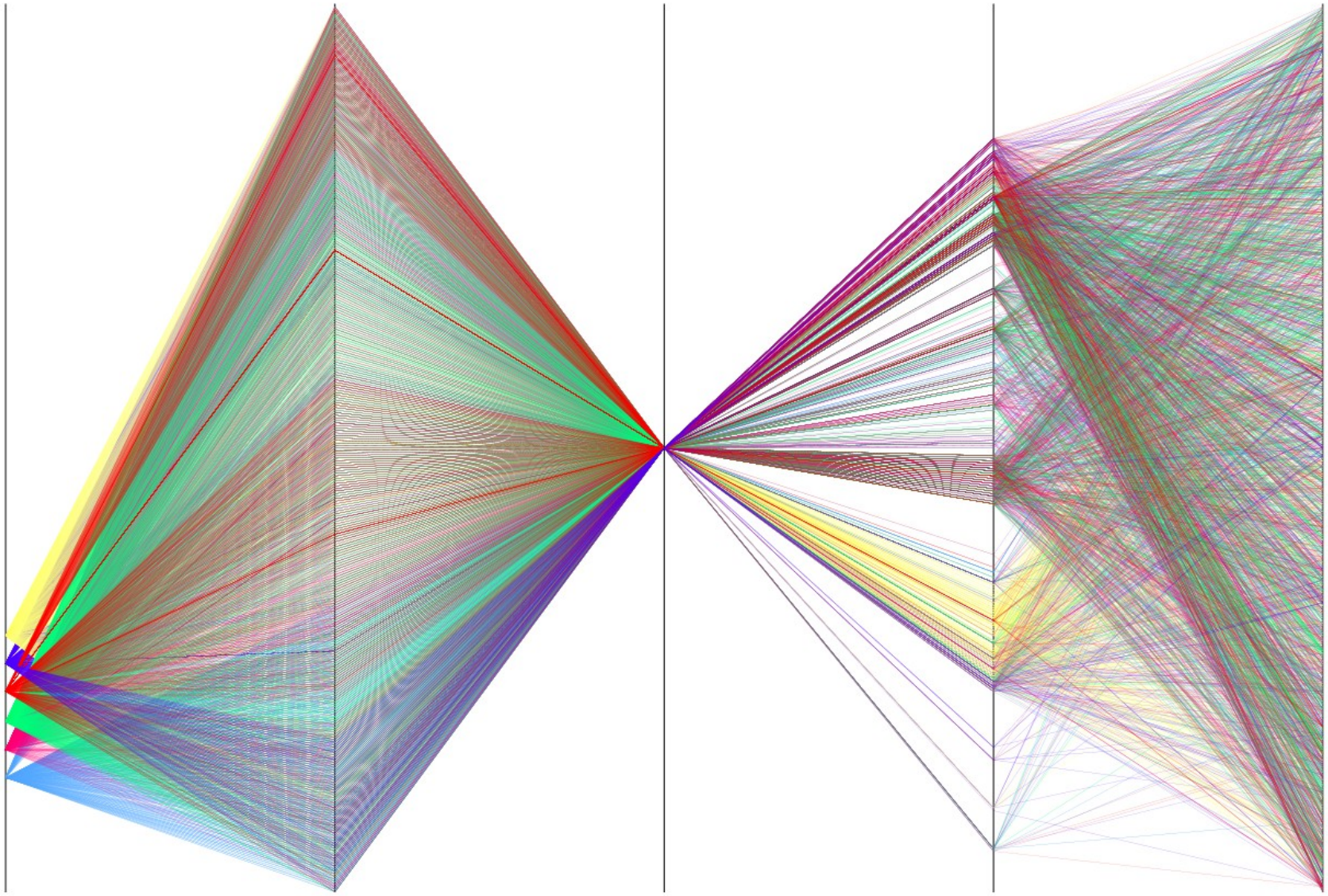
Análise sshd - PicViz

- Dias: 4, 5, 6, 7 e 8.



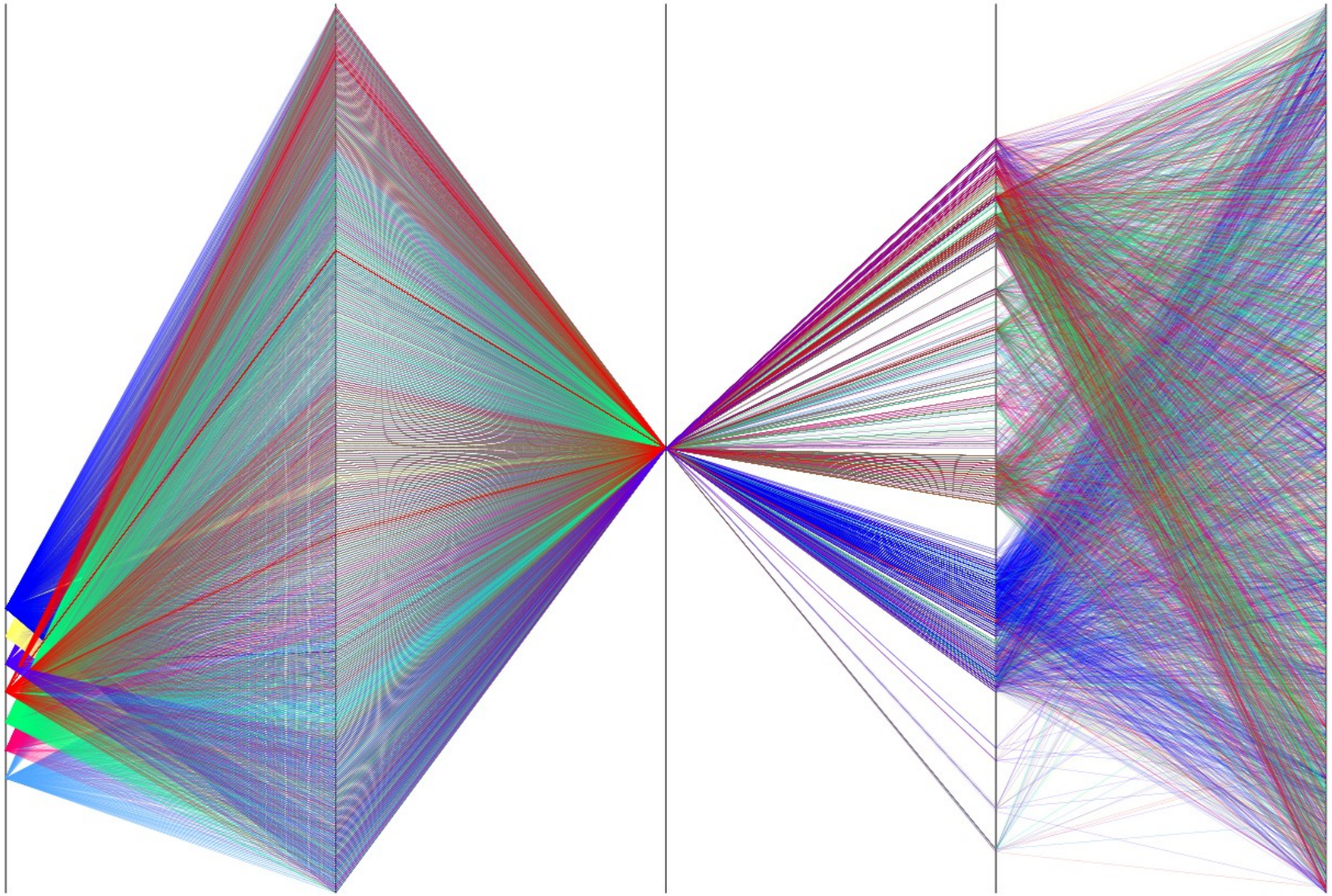
Análise sshd - PicViz

- Dias: 4, 5, 6, 7, 8 e 9.



Análise sshd - PicViz

- Dias: 4, 5, 6, 7, 8, 9 e 10.



- Invocando com filtro:

  # picviz-gui dados.pgdl 'hide value !="200.*" on axis 4'

  - Isolei somente dias 8,9 e 10.
  - Mudança no comportamento do eixo com os Ips (senão não adiantaria usar o filtro).

Conclusões

- Scan distribuído de forma inteligente;
- Olhe novamente o seu *denyhosts*, e considere 3 logins falhos em pelo menos 1 semana;
- Enquanto isso na web...
  - http://isc.sans.edu/diary.html?storyid=9370
  - http://www.computersecurityarticles.info/tag/force/

Bye!

- Obrigado!

gabriel@las.ic.unicamp.br

Ps: Me dá seu *log*??