mix_protocols

Rafilx

2022-06-12

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Loading required package: viridisLite

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

R Markdown

Analisar a porcentagem de ataques/requisições por protocolo, dividindo por períodos. Essa análise não envolve os payloads, apenas os quantitativos de ataques e requisições.

Resultados esperados:

- gráficos de linhas e de barras mostrando a evolução (ver ideias na planilha)

```
db <- dbConnect(RSQLite::SQLite(), dbname="../db/database-2022-05-11/mix_protocol.sqlite")

data_unfetch <-dbSendQuery(db, "
  SELECT *, CAST(CAST(year AS text) || CAST(period AS text) as integer) as year_period
    FROM (
      SELECT *,  strftime(\"%Y\", tempo_inicio) as year, ((strftime(\"%m\", tempo_final) - 1) / 3) + 1
        FROM MIX_PROTOCOL
    )
")
data <- fetch(data_unfetch)

dbDisconnect(db)
```

```
## Warning in connection_release(conn@ptr): There are 1 result in use. The
## connection will be released when they are closed
```

- Agrupamento realizado por período (trimestre) e "attack_protocol" é o protocolo utilizado no ataque `["chargen", "cldap", "coap", "dns", "memcached", "ntp", "qotd", "ssdp", "steam_games"]`

```
data['tempo_final_cast'] = as.POSIXct(data[['tempo_final']], format = "%Y-%m-%d %H:%M:%S")
data['tempo_inicio_cast'] = as.POSIXct(data[['tempo_inicio']], format = "%Y-%m-%d %H:%M:%S")

data_grouped_period_protocol = data %>%
  mutate(year_period_int = year_period,
         year_period = as.factor(year_period),
         attack_protocol = as.factor(attack_protocol)) %>%
  group_by(year_period, attack_protocol) %>%
  summarise(sum_requests_per_attack = sum(requests_per_attack),
            number_of_attacks = n(),
            tempo_inicio=min(tempo_inicio_cast),
            tempo_final=max(tempo_final_cast))
```

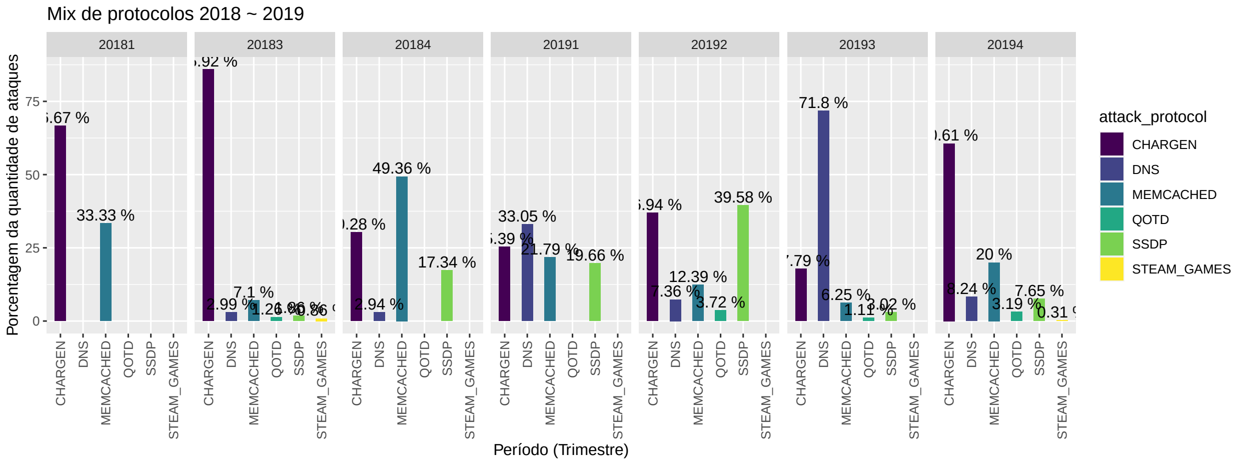
```
## ‘summarise()‘ has grouped output by ’year_period’. You can override using the
## ‘.groups‘ argument.
```

```
data_grouped_period_protocol_percentage = data_grouped_period_protocol %>%
  ungroup() %>%
  group_by(year_period) %>%
  summarise(attack_protocol = attack_protocol,
            number_of_attacks = number_of_attacks,
            tempo_inicio = tempo_inicio,
            tempo_final = tempo_final,
            sum_period_number_of_attacks = sum(number_of_attacks),
            sum_period_requests_per_attack = sum(sum_requests_per_attack),
            sum_requests_per_attack = sum_requests_per_attack) %>%
  mutate(number_of_attacks_percentage = (number_of_attacks / sum_period_number_of_attacks) * 100,
         number_of_requests_percentage = (sum_requests_per_attack / sum_period_requests_per_attack) * 10
```

```
## ‘summarise()‘ has grouped output by ’year_period’. You can override using the
## ‘.groups‘ argument.
```

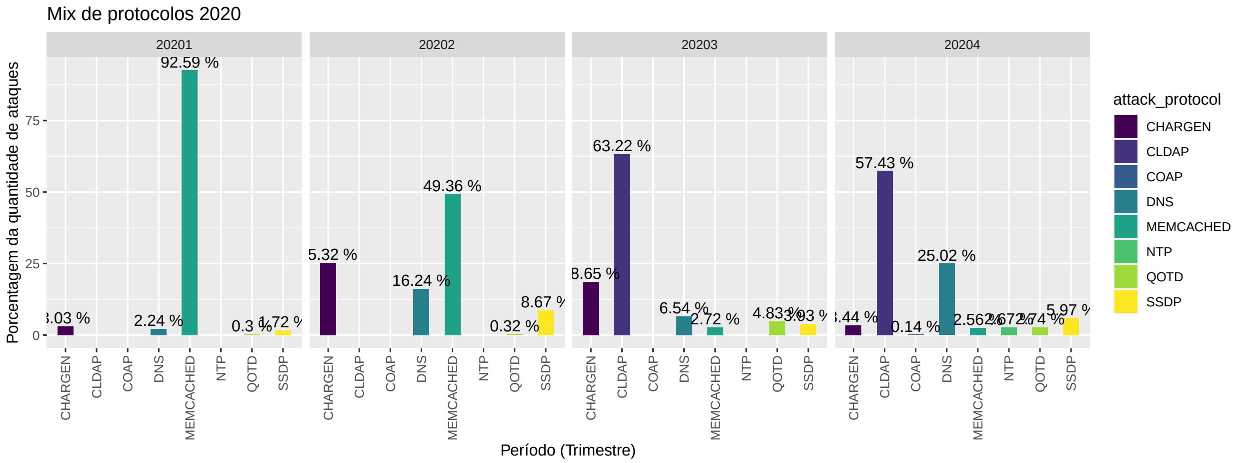
- Gráfico de barras 2018 e 2019

```
data_grouped_period_protocol_percentage %>%
  filter(year_period  %in% c(20181, 20182, 20183, 20184, 20191, 20192, 20193, 20194)) %>%
  filter(number_of_attacks_percentage > 0.1) %>%
  ggplot( aes(x=attack_protocol, y=number_of_attacks_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5, position="dodge") +
    geom_text(aes(label = paste(round(number_of_attacks_percentage, 2), "%"),  vjust = -0.25)) +
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    facet_grid(~year_period) +
    ylab("Porcentagem da quantidade de ataques") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2018 ~ 2019")
```



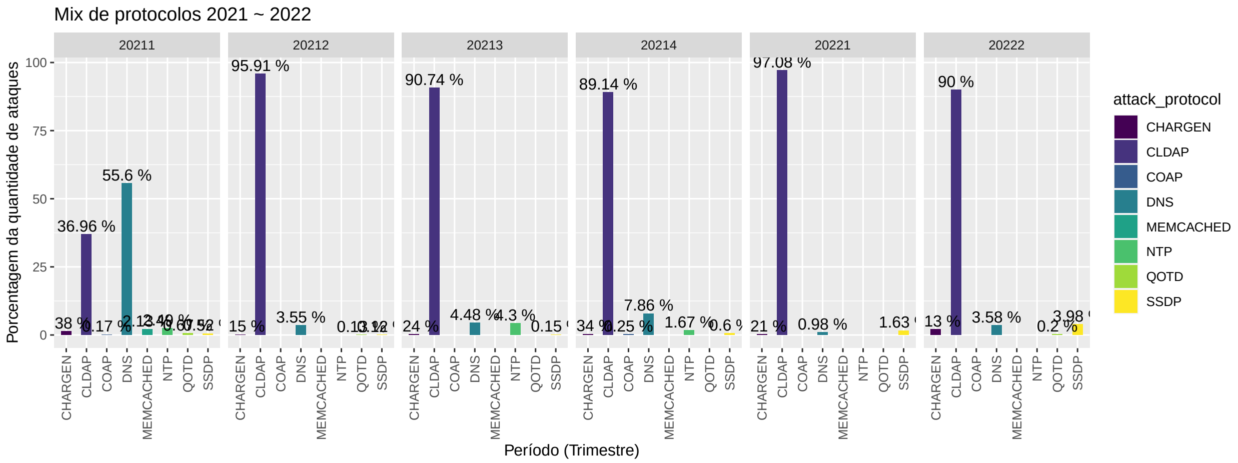
- Gráfico de barras 2020

```
data_grouped_period_protocol_percentage %>%
  filter(year_period  %in% c(20201, 20202, 20203, 20204)) %>%
  filter(number_of_attacks_percentage > 0.1) %>%
  ggplot( aes(x=attack_protocol, y=number_of_attacks_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5, position="dodge") +
    geom_text(aes(label = paste(round(number_of_attacks_percentage, 2), "%"),  vjust = -0.25)) +
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    facet_grid(~year_period) +
    ylab("Porcentagem da quantidade de ataques") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2020")
```



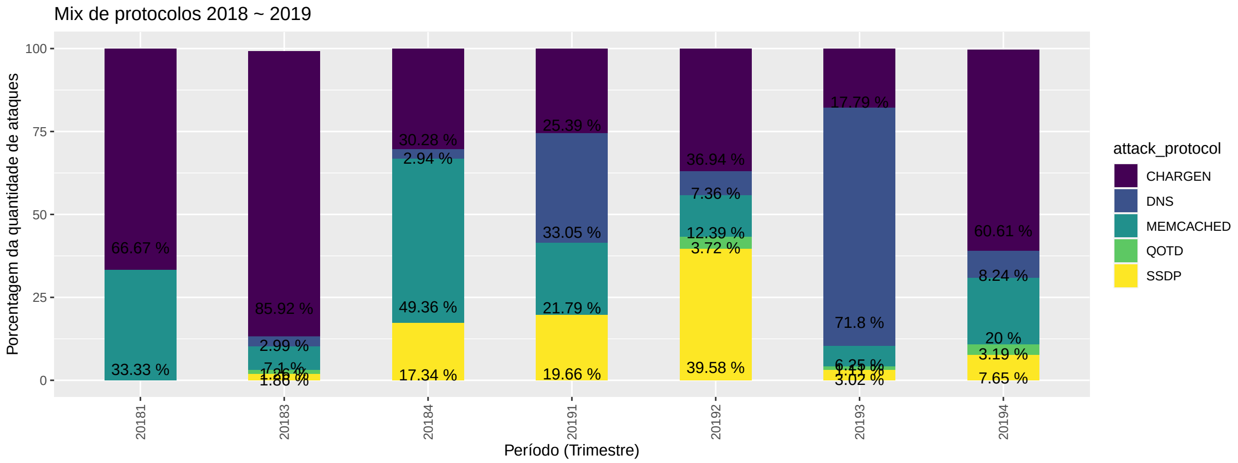
- Gráfico de barras 2021 ~ 2022

```
data_grouped_period_protocol_percentage %>%
  filter(year_period  %in% c(20211, 20212, 20213, 20214, 20221, 20222, 20223)) %>%
  filter(number_of_attacks_percentage > 0.1) %>%
  ggplot( aes(x=attack_protocol, y=number_of_attacks_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5, position="dodge") +
    geom_text(aes(label = paste(round(number_of_attacks_percentage, 2), "%"),  vjust = -0.25)) +
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    facet_grid(~year_period) +
    ylab("Porcentagem da quantidade de ataques") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2021 ~ 2022")
```



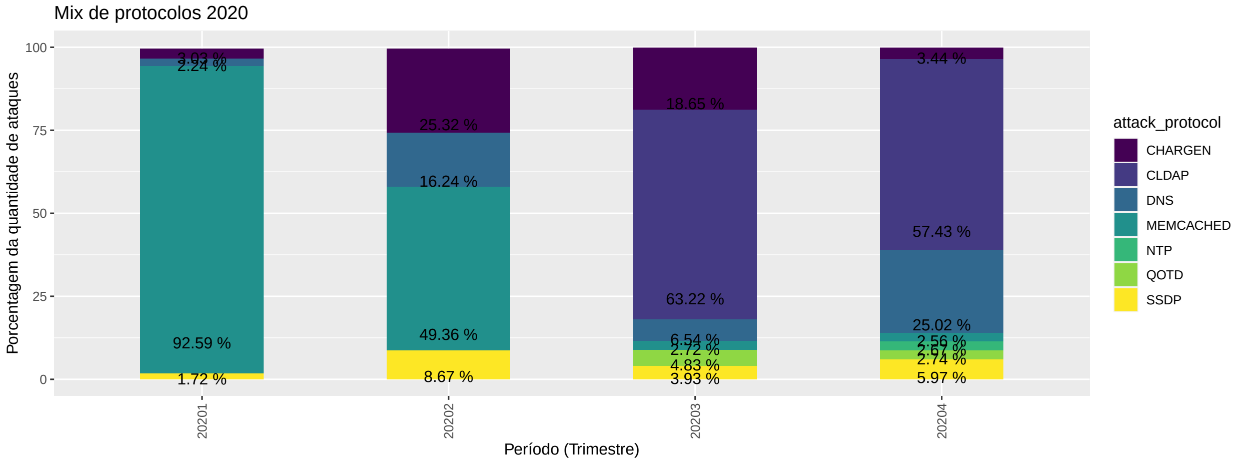
- Gráfico de barras empilhadas 2018 ~ 2019

```
data_grouped_period_protocol_percentage %>%
  filter(year_period  %in% c(20181, 20182, 20183, 20184, 20191, 20192, 20193, 20194)) %>%
  filter(number_of_attacks_percentage > 1) %>%
  ggplot( aes(x=year_period, y=number_of_attacks_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5) +
    geom_text(aes(label = paste(round(number_of_attacks_percentage, 2), "%")), position = position_stac
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    ylab("Porcentagem da quantidade de ataques") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2018 ~ 2019")
```



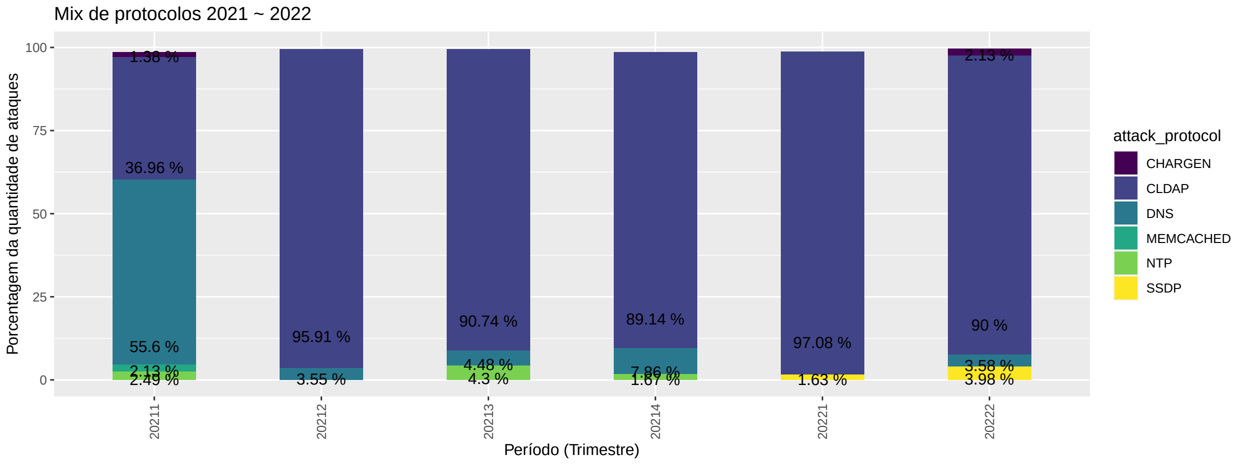
- Gráfico de barras empilhadas 2020

```
data_grouped_period_protocol_percentage %>%
  filter(year_period  %in% c(20201, 20202, 20203, 20204)) %>%
  filter(number_of_attacks_percentage > 1) %>%
  ggplot( aes(x=year_period, y=number_of_attacks_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5) +
    geom_text(aes(label = paste(round(number_of_attacks_percentage, 2), "%")), position = position_stac
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    ylab("Porcentagem da quantidade de ataques") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2020")
```



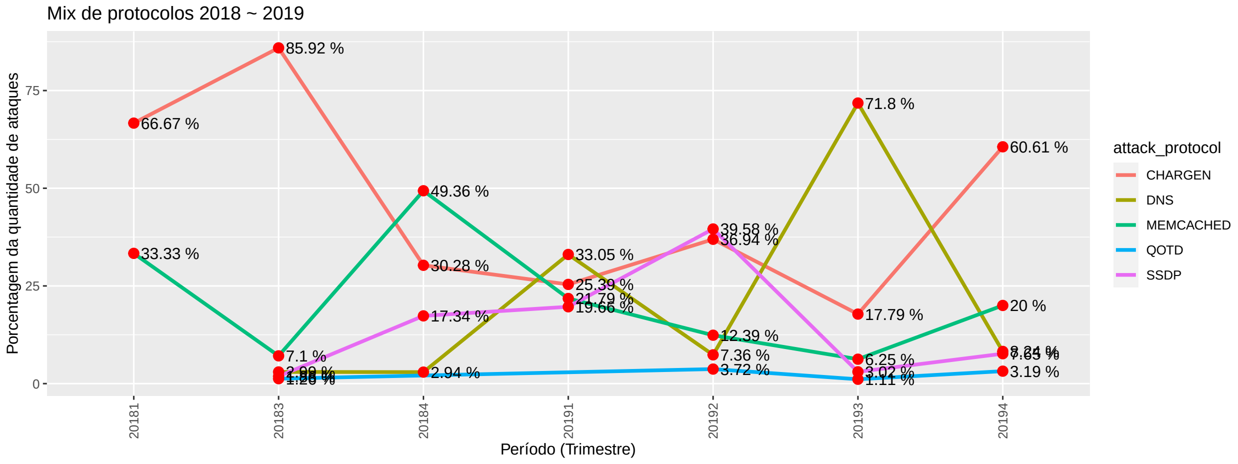
- Gráfico de barras empilhadas 2021 ~ 2022

```
data_grouped_period_protocol_percentage %>%
  filter(year_period  %in% c(20211, 20212, 20213, 20214, 20221, 20222, 20223)) %>%
  filter(number_of_attacks_percentage > 1) %>%
  ggplot( aes(x=year_period, y=number_of_attacks_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5) +
    geom_text(aes(label = paste(round(number_of_attacks_percentage, 2), "%")), position = position_stac
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    ylab("Porcentagem da quantidade de ataques") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2021 ~ 2022")
```



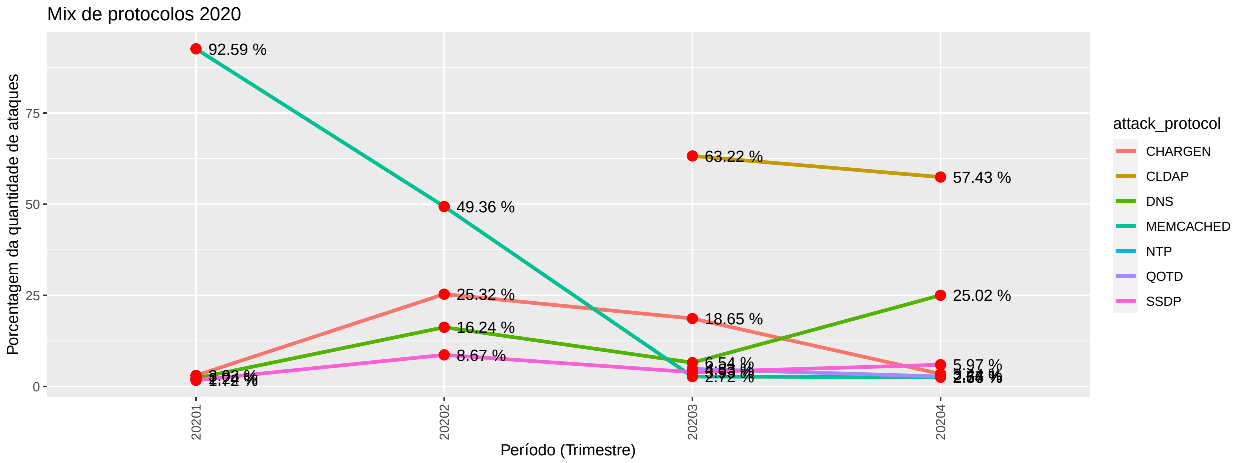
- Gráfico de linhas 2018 ~ 2019

```
data_grouped_period_protocol_percentage %>%
  filter(year_period  %in% c(20181, 20182, 20183, 20184, 20191, 20192, 20193, 20194)) %>%
  filter(number_of_attacks_percentage > 1) %>%
  ggplot( aes(x=year_period, y=number_of_attacks_percentage, group=attack_protocol)) +
    geom_line(size=1.2, aes(color=attack_protocol)) +
    geom_point(color="red", size=3,  aes(color=attack_protocol)) +
    geom_text(
      aes(label = paste(round(number_of_attacks_percentage, 2), "%")),
      hjust = 0, nudge_x = 0.05,
    ) +
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    ylab("Porcentagem da quantidade de ataques") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2018 ~ 2019")
```



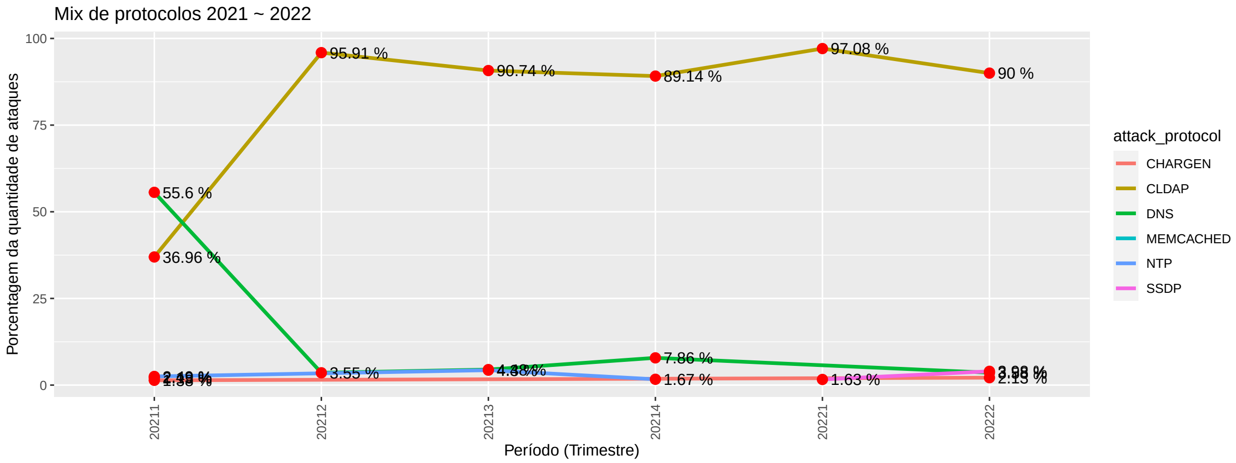
- Gráfico de linhas 2020

```r
data_grouped_period_protocol_percentage %>%
  filter(year_period  %in% c(20201, 20202, 20203, 20204)) %>%
  filter(number_of_attacks_percentage > 1) %>%
  ggplot( aes(x=year_period, y=number_of_attacks_percentage, group=attack_protocol)) +
    geom_line(size=1.2, aes(color=attack_protocol)) +
    geom_point(color="red", size=3,  aes(color=attack_protocol)) +
    geom_text(
      aes(label = paste(round(number_of_attacks_percentage, 2), "%")),
      hjust = 0, nudge_x = 0.05,
    ) +
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    ylab("Porcentagem da quantidade de ataques") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2020")
```



- Gráfico de linhas 2021 ~ 2022

```r
data_grouped_period_protocol_percentage %>%
  filter(year_period  %in% c(20211, 20212, 20213, 20214, 20221, 20222, 20223)) %>%
  filter(number_of_attacks_percentage > 1) %>%
  ggplot( aes(x=year_period, y=number_of_attacks_percentage, group=attack_protocol)) +
    geom_line(size=1.2, aes(color=attack_protocol)) +
    geom_point(color="red", size=3,  aes(color=attack_protocol)) +
    geom_text(
      aes(label = paste(round(number_of_attacks_percentage, 2), "%")),
      hjust = 0, nudge_x = 0.05,
    ) +
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    ylab("Porcentagem da quantidade de ataques") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2021 ~ 2022")
```


Mix de protocolos 2021 ~ 2022
Porcentagem da quantidade de ataques
100
75
50
25
0
95.91 %
90.74 %
89.14 %
97.08 %
90 %
55.6 %
36.96 %
3.55 %
7.86 %
1.67 %
1.63 %
2.15 %
20211
20212
20213
20214
20221
20222
Período (Trimestre)
attack_protocol
CHARGEN
CLDAP
DNS
MEMCACHED
NTP
SSDP