memcached analysis

Rafilx

2022-06-19

```
##
## 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: gridExtra

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine

## Loading required package: viridisLite

##
## Attaching package: 'lubridate'

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

R Markdown

Na dissertação, foi relatado que

- 3,1% das requisições de Memcached eram variações de `stats`
- 91,6% das requisições eram `set` ou `get`
- 5,1% das requisições eram malformadas (e.g., requisições HTTP ou SSDP) ou eram `flush_all` (para limpar chaves do cache)

É possível fazer essa análise por trimestre, ou ao menos analisar a incidência de `stats` e `set`+`get`

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


data_unfetch <-dbSendQuery(db, "
  SELECT ip, requests_per_attack, tempo_inicio, tempo_final, QUOTE(payload) as quote_payload, QUOTE(get_
         payload_decoded, payload, raw_payload,
         CAST(CAST(year AS text) || CAST(period AS text) as integer) as year_period, SUBSTR(payload,0,25
    FROM (
      SELECT *,  strftime(\"%Y\", tempo_inicio) as year, ((strftime(\"%m\", tempo_final) - 1) / 3) + 1 A
        FROM MEMCACHED_ANALYSIS
    )
")

data <- fetch(data_unfetch)

data_memcached_payload_types_unfetch <-dbSendQuery(db, "
  SELECT id, quantity, SUBSTR(payload,0,25) AS payload_limit
    FROM MEMCACHED_PAYLOAD_TYPES
")
```

```
## Warning: Closing open result set, pending rows
```

```
data_memcached_payload_types <- fetch(data_memcached_payload_types_unfetch)


dbDisconnect(db)
```

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

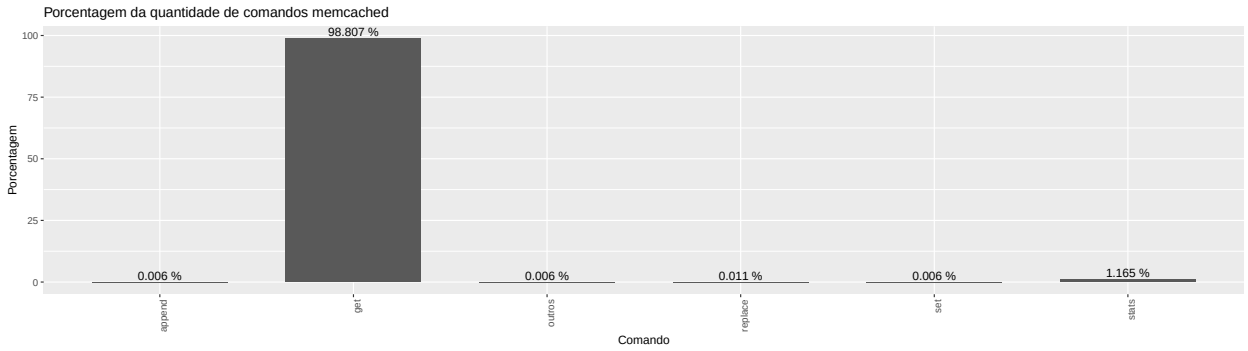
```
data['tempo_final_cast'] = as.POSIXct(data[['tempo_final']], format = "%Y-%m-%d %H:%M:%S")
data['tempo_inicio_cast'] = as.POSIXct(data[['tempo_inicio']], format = "%Y-%m-%d %H:%M:%S")
```

```
memcached_payload_types = data_memcached_payload_types %>%
  mutate(payload_str = toString(payload_limit)) %>%
  arrange(desc(quantity)) %>%
  select('quantity', 'payload_limit', 'id')

memcached_payload_types_quantity_percentage = memcached_payload_types %>%
  mutate(sum_quantity = sum(quantity)) %>%
  mutate(quantity_percentage = (quantity / sum_quantity) * 100)

memcached_payload_types_quantity_percentage %>%
  select('quantity', 'quantity_percentage', 'payload_limit') %>%
  arrange(desc(quantity)) %>%
  head(15)
```

```
##    quantity quantity_percentage payload_limit
## 1     17719           98.806669           get
## 2       209            1.165449         stats
## 3         2            0.011153       replace
```

```
## 4          1            0.005576           set
## 5          1            0.005576        append
## 6          1            0.005576        outros
## 7          0            0.000000           add
## 8          0            0.000000           cas
## 9          0            0.000000       prepend
## 10         0            0.000000     flush_all
```

```
memcached_payload_types_quantity_percentage %>%
 arrange(desc(quantity)) %>%
 filter(quantity > 0) %>%
 select('quantity_percentage', 'payload_limit') %>%
 ggplot( aes(x=payload_limit, y=quantity_percentage)) +
   geom_bar(stat="identity", width = 0.7, position="dodge") +
   geom_text(aes(label = paste(round(quantity_percentage, 3), "%"),  vjust = -0.25)) +
   scale_fill_viridis(discrete=TRUE, direction = -1) +
   theme(axis.text.x = element_text(angle = 90, vjust = 1, hjust=1)) +
   ylab("Porcentagem") +
   xlab("Comando") +
   ggtitle("Porcentagem da quantidade de comandos memcached")
```



- Agrupamento realizado por período (trimestre) e "memcached_request_type" é o comando utilizado no ataque `["stats", "set", "get", "add", "cas", "replace", "append", "prepend", "flush_all", "outros"]`
- Somando a quantidade de requisições utilizadas por cada comando e período

```
data_grouped_period_command = data %>%
  mutate(year_period_int = year_period,
         year_period = as.factor(year_period),
         command = as.factor(memcached_request_type)) %>%
  group_by(year_period, command) %>%
  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_command_percentage = data_grouped_period_command %>%
  ungroup() %>%
  group_by(year_period) %>%
  summarise(command = command,
            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) * 1
```

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

```
minimum_percentage_as_others = 1
decimals_digits = 1

data_grouped_period_command_others_percentage = data_grouped_period_command_percentage %>%
  mutate(
    command = case_when(
      number_of_requests_percentage < minimum_percentage_as_others ~ "OUTROS",
      TRUE ~ as.character(command)
    )
  ) %>%
  group_by(year_period, command) %>%
  summarise(number_of_requests_percentage = sum(number_of_requests_percentage))
```

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

```
data_grouped_period_command_others_percentage %>%
  print(n=16)
```

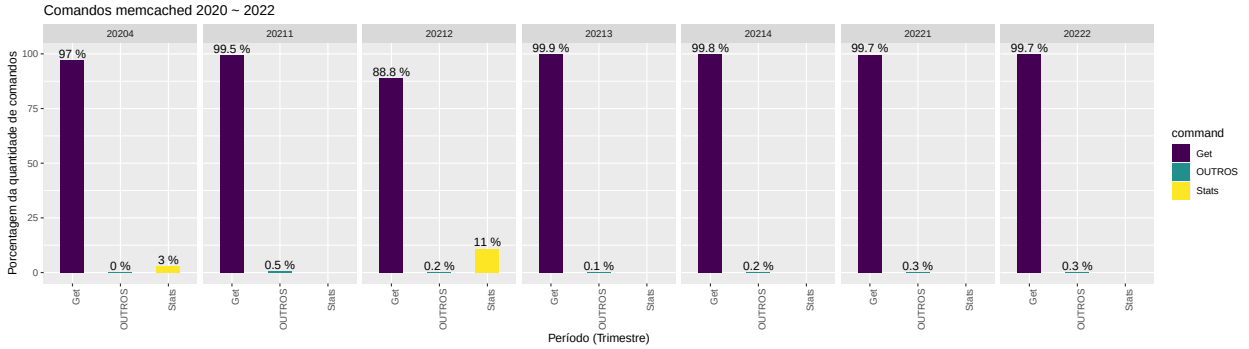
```
## # A tibble: 16 x 3
## # Groups:   year_period [7]
##    year_period command number_of_requests_percentage
##    <fct>       <chr>                           <dbl>
##  1 20204       Get                          97.0
##  2 20204       OUTROS                        0.00000656
##  3 20204       Stats                         3.00
##  4 20211       Get                          99.5
##  5 20211       OUTROS                        0.466
##  6 20212       Get                          88.8
##  7 20212       OUTROS                        0.220
##  8 20212       Stats                        11.0
##  9 20213       Get                          99.9
## 10 20213       OUTROS                        0.0763
## 11 20214       Get                          99.8
## 12 20214       OUTROS                        0.211
## 13 20221       Get                          99.7
```

```
## 14 20221        OUTROS                    0.308
## 15 20222        Get                      99.7
## 16 20222        OUTROS                    0.290
```

Porcentagem menores que 1 foram agrupadas como “OUTROS”

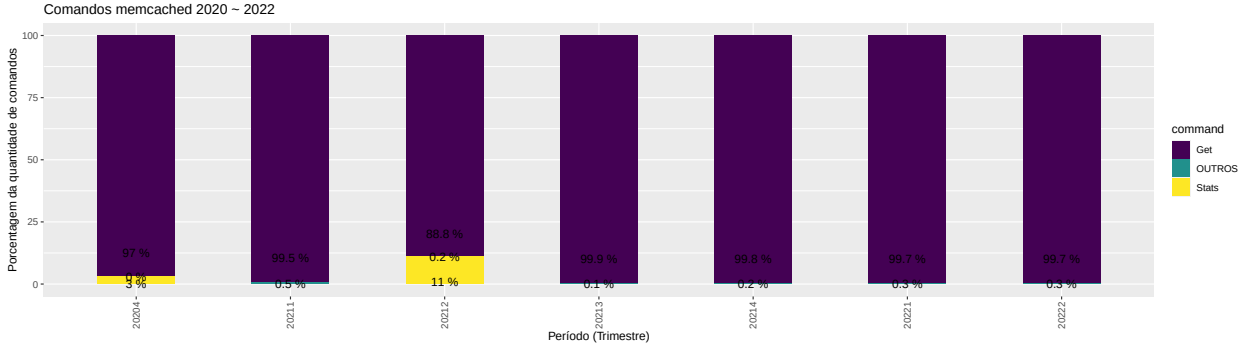
- Gráfico de barras 2020 ~ 2022

```
data_grouped_period_command_others_percentage %>%
  ggplot( aes(x=command, y=number_of_requests_percentage, fill=command)) +
    geom_bar(stat="identity", width = 0.5, position="dodge") +
    geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%"),  vjust = -0
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 1, hjust=1)) +
    facet_grid(~year_period) +
    ylab("Porcentagem da quantidade de comandos") +
    xlab("Período (Trimestre)") +
    ggtitle("Comandos memcached 2020 ~ 2022")
```



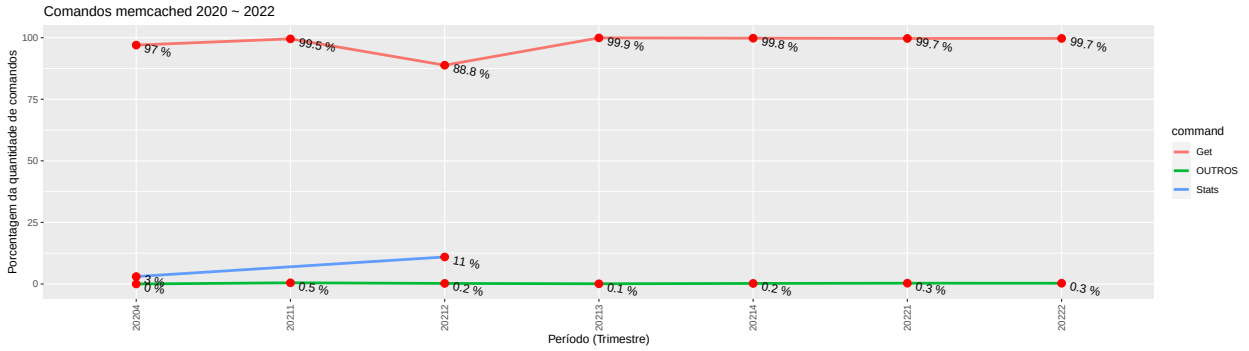
- Gráfico de barras empilhadas 2020 ~ 2022

```
data_grouped_period_command_others_percentage %>%
  ggplot( aes(x=year_period, y=number_of_requests_percentage, fill=command)) +
    geom_bar(stat="identity", width = 0.5) +
    geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%")), position =
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    ylab("Porcentagem da quantidade de comandos") +
    xlab("Período (Trimestre)") +
    ggtitle("Comandos memcached 2020 ~ 2022")
```



- Gráfico de linhas 2020 ~ 2022

```
data_grouped_period_command_others_percentage %>%
  ggplot( aes(x=year_period, y=number_of_requests_percentage, group=command)) +
    geom_line(size=1.2, aes(color=command)) +
    geom_point(color="red", size=3,  aes(color=command)) +
    geom_text(
      aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%")),
      hjust = -0.03, nudge_x = 0.05, nudge_y = -1, angle = -10,
    ) +
    scale_fill_viridis(discrete=TRUE) +
    theme(
      axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
    ) +
    ylab("Porcentagem da quantidade de comandos") +
    xlab("Período (Trimestre)") +
    ggtitle("Comandos memcached 2020 ~ 2022")
```




```
data_grouped_period_payload = data %>%
  mutate(year_period = as.factor(year_period),
         payload = as.factor(raw_payload)) %>%
  group_by(year_period, payload) %>%
  summarise(sum_requests_per_attack = sum(requests_per_attack),
            number_of_payloads_by_period = n())
```

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

```
data_grouped_period_payload_percentage = data_grouped_period_payload %>%
  ungroup() %>%
  group_by(year_period) %>%
  summarise(payload = payload,
            number_of_payloads_by_period = number_of_payloads_by_period,
            sum_number_of_payloads_by_period = sum(number_of_payloads_by_period),
            sum_period_requests_per_attack = sum(sum_requests_per_attack),
            sum_requests_per_attack = sum_requests_per_attack) %>%
  mutate(number_of_payloads_percentage = (number_of_payloads_by_period / sum_number_of_payloads_by_peri
         number_of_requests_percentage = (sum_requests_per_attack / sum_period_requests_per_attack) * 1
```

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

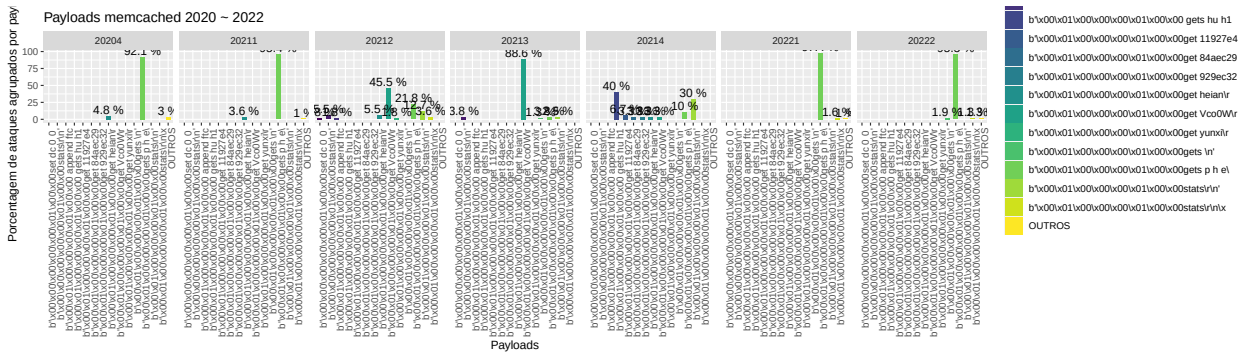
```
data_grouped_period_payload_others_percentage = data_grouped_period_payload_percentage %>%
  mutate(
    payload = case_when(
      number_of_payloads_percentage < minimum_percentage_as_others ~ "OUTROS",
      TRUE ~ as.character(payload)
    )
  ) %>%
  group_by(year_period, payload) %>%
  summarise(
    number_of_payloads_percentage = sum(number_of_payloads_percentage))
```

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

Porcentagem menores que 1 foram agrupadas como "OUTROS" payload foi limitado a mostrar os 45 primeiros characteres

- Esses dados podem ser analisados como
  - year_period = ano e trimestre
  - payload = o payload utilizado no ataque, aqueles payloads pouco utilizados > 1 foi agrupado como "OUTROS"
  - number_of_payloads_percentage = porcentagem da quantidade de ataques que utilizaram o mesmo payload naquele período nesse caso cada trimestre (year_period) tem um somatório de 100% do campo (number_of_payloads_percentage)
- Gráfico de barras 2020 ~ 2022

```
data_grouped_period_payload_others_percentage %>%
  mutate(payload_limit=substr(payload, 0, 45)) %>%
  ggplot( aes(x=payload_limit, y=number_of_payloads_percentage, fill=payload_limit)) +
    geom_bar(stat="identity", width = 0.5, position="dodge") +
    geom_text(aes(label = paste(round(number_of_payloads_percentage, decimals_digits), "%"),  vjust = -0
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 1, hjust=1)) +
    facet_grid(~year_period) +
    ylab("Porcentagem de ataques agrupados por payload e periodo") +
    xlab("Payloads") +
    ggtitle("Payloads memcached 2020 ~ 2022")
```



```
data_grouped_period_payload_others_percentage %>%
  mutate(
    payload_limit=case_when(
      payload == "OUTROS" ~ as.character(payload),
      TRUE ~ substr(payload, 35, 80)
    )
    ) %>%
  select('year_period', 'percentage'='number_of_payloads_percentage', 'payload_limit') %>%
  print(n=35)
```

```
## # A tibble: 35 x 3
## # Groups:   year_period [7]
##    year_period percentage payload_limit
##    <fct>            <dbl> <chr>
##  1 20204            4.82  "get heian\\r\\n'"
##  2 20204           92.1   "gets p h e\\n'"
##  3 20204            3.04  "OUTROS"
##  4 20211            3.58  "get heian\\r\\n'"
##  5 20211           95.4   "gets p h e\\n'"
##  6 20211            0.982 "OUTROS"
##  7 20212            1.82  "set dc 0 0 1273 noreply\\r\\nAtatata taata taata"
##  8 20212            5.45  "stats\\r\\n'"
##  9 20212            1.82  " append ftca 0 864000 150\\r\\n_PAY_0.1_BTC_TO_3"
## 10 20212            5.45  "get heian\\r\\n'"
## 11 20212           45.5   "get VcoOW\\r\\nget VcoOW\\r\\nget VcoOW\\r\\nget Vco"
## 12 20212            1.82  "get yunxi\\r\\n'"
## 13 20212           21.8   "gets p h e\\n'"
## 14 20212           12.7   "stats\\r\\n'"
## 15 20212            3.64  "stats\\r\\n\\x00'"
## 16 20213            3.80  "stats\\r\\n'"
## 17 20213           88.6   "get VcoOW\\r\\nget VcoOW\\r\\nget VcoOW\\r\\nget Vco"
## 18 20213            1.27  "gets \\n'"
## 19 20213            3.80  "gets p h e\\n'"
## 20 20213            2.53  "stats\\r\\n'"
## 21 20214           40     " gets hu h1 hua hub huc hud hue\\r\\n\\x00'"
## 22 20214            6.67  "get 11927e4abe45a174b3f3b5b8ea823076\\r\\n'"
## 23 20214            3.33  "get 84aec29299912348b585ad2d30de290a\\r\\n'"
## 24 20214            3.33  "get 929ec32cd07d4f0da08562c1ab55c8df\\r\\n'"
## 25 20214            3.33  "get heian\\r\\n'"
## 26 20214            3.33  "get VcoOW\\r\\nget VcoOW\\r\\nget VcoOW\\r\\nget Vco"
## 27 20214           10     "gets p h e\\n'"
## 28 20214           30     "stats\\r\\n'"
## 29 20221           97.4   "gets p h e\\n'"
## 30 20221            1.62  "stats\\r\\n\\x00'"
## 31 20221            0.971 "OUTROS"
## 32 20222            1.94  "gets \\n'"
## 33 20222           95.5   "gets p h e\\n'"
## 34 20222            1.29  "stats\\r\\n\\x00'"
## 35 20222            1.29  "OUTROS"
```

Dados do payload get

- São o resto do payload após o "get"

```
data_get_payload = data %>%
  filter(memcached_request_type %in% c("Get")) %>%
  select(year_period, requests_per_attack, quote_payload_get, raw_payload)

data_grouped_period_payload_get = data_get_payload %>%
  mutate(year_period = as.factor(year_period),
         payload_get = as.factor(quote_payload_get)) %>%
  group_by(year_period, payload_get) %>%
  summarise(sum_requests_per_attack = sum(requests_per_attack),
            number_of_payloads_get_by_period = n())
```

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

```
data_grouped_period_payload_get_percentage = data_grouped_period_payload_get %>%
  ungroup() %>%
  group_by(year_period) %>%
  summarise(payload_get = payload_get,
            number_of_payloads_get_by_period = number_of_payloads_get_by_period,
            sum_number_of_payloads_get_by_period = sum(number_of_payloads_get_by_period),
            sum_period_requests_per_attack = sum(sum_requests_per_attack),
            sum_requests_per_attack = sum_requests_per_attack) %>%
  mutate(number_of_payloads_get_percentage = (number_of_payloads_get_by_period / sum_number_of_payloads
         number_of_requests_percentage = (sum_requests_per_attack / sum_period_requests_per_attack) * 1
```

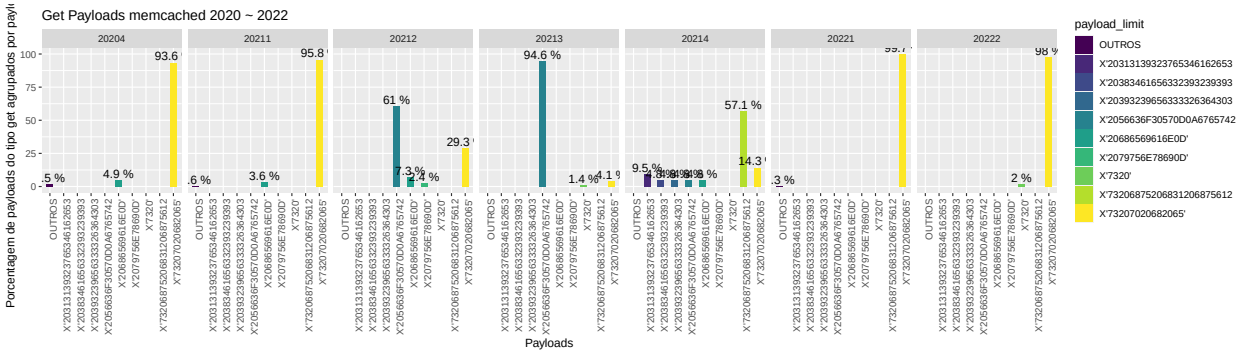
```
## `summarise()` has grouped output by 'year_period'. You can override using the
## `.groups` argument.
```

```
data_grouped_period_payload_get_others_percentage = data_grouped_period_payload_get_percentage %>%
  mutate(
    payload_get = case_when(
      number_of_payloads_get_percentage < minimum_percentage_as_others ~ "OUTROS",
      TRUE ~ as.character(payload_get)
    )
  ) %>%
  group_by(year_period, payload_get) %>%
  summarise(
    number_of_payloads_get_percentage = sum(number_of_payloads_get_percentage)
    )
```

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

- Esse gráfico não ficou bacana, mas da uma noção
- Nesses dados, foi pego somente os registros que eram do tipo "get" e os payloads após o byte 'get', que tecnicamente representa o que foi a key buscada pelo comando get
- Gráfico de barras 2020 ~ 2022

```
data_grouped_period_payload_get_others_percentage %>%
  mutate(payload_limit=substr(payload_get, 0, 25)) %>%
  ggplot( aes(x=payload_limit, y=number_of_payloads_get_percentage, fill=payload_limit)) +
    geom_bar(stat="identity", width = 0.5, position="dodge") +
    geom_text(aes(label = paste(round(number_of_payloads_get_percentage, decimals_digits), "%"),  vjust
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 1, hjust=1)) +
    facet_grid(~year_period) +
    ylab("Porcentagem de payloads do tipo get agrupados por payload e periodo") +
    xlab("Payloads") +
    ggtitle("Get Payloads memcached 2020 ~ 2022")
```




```
data_grouped_period_payload_get_others_percentage_with_raw_payload = left_join(x=data_grouped_period_pay
          by=c("payload_get" = "quote_payload_get"),
          keep=FALSE,
          copy = FALSE
          ) %>%
  count(year_period.x, payload_get, number_of_payloads_get_percentage, raw_payload) %>%
  select(year_period = year_period.x, payload_get,  percentage=number_of_payloads_get_percentage, raw_pa


data_grouped_period_payload_get_others_percentage_with_raw_payload %>%
  mutate(
    payload_limit=case_when(
      payload_get == "OUTROS" ~ as.character(payload_get),
      TRUE ~ substr(raw_payload, 35, 80)
    )
    ) %>%
  select(year_period, percentage, payload_limit) %>%
  print(n=24)
```

```
## # A tibble: 24 x 3
##    year_period percentage payload_limit
##    <fct>            <dbl> <chr>
##  1 20204             1.52  "OUTROS"
##  2 20204             4.89  "get heian\\r\\n'"
##  3 20204            93.6   "gets p h e\\n'"
##  4 20211             0.635 "OUTROS"
##  5 20211             3.59  "get heian\\r\\n'"
##  6 20211            95.8   "gets p h e\\n'"
##  7 20212            61.0   "get VcoOW\\r\\nget VcoOW\\r\\nget VcoOW\\r\\nget Vco"
```

```
##  8 20212           7.32  "get heian\\r\\n'"
##  9 20212           2.44  "get yunxi\\r\\n'"
## 10 20212          29.3   "gets p h e\\n'"
## 11 20213          94.6   "get VcoOW\\r\\nget VcoOW\\r\\nget VcoOW\\r\\nget Vco"
## 12 20213           1.35  "gets \\n'"
## 13 20213           4.05  "gets p h e\\n'"
## 14 20214           9.52  "get 11927e4abe45a174b3f3b5b8ea823076\\r\\n'"
## 15 20214           4.76  "get 84aec29299912348b585ad2d30de290a\\r\\n'"
## 16 20214           4.76  "get 929ec32cd07d4f0da08562c1ab55c8df\\r\\n'"
## 17 20214           4.76  "get VcoOW\\r\\nget VcoOW\\r\\nget VcoOW\\r\\nget Vco"
## 18 20214           4.76  "get heian\\r\\n'"
## 19 20214          57.1   " gets hu h1 hua hub huc hud hue\\r\\n\\x00'"
## 20 20214          14.3   "gets p h e\\n'"
## 21 20221           0.331 "OUTROS"
## 22 20221          99.7   "gets p h e\\n'"
## 23 20222           1.99  "gets \\n'"
## 24 20222          98.0   "gets p h e\\n'"
```