# Less Configuration, More Security:

## Automated discovery of asset- and user roles

Stanislav Miskovic, PhD, Principal Data Scientist
Prasoon Shukla, Sr. Product Manager
Dimitrios Terzis, PhD, Principal Software Engineer

October 2018

# Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

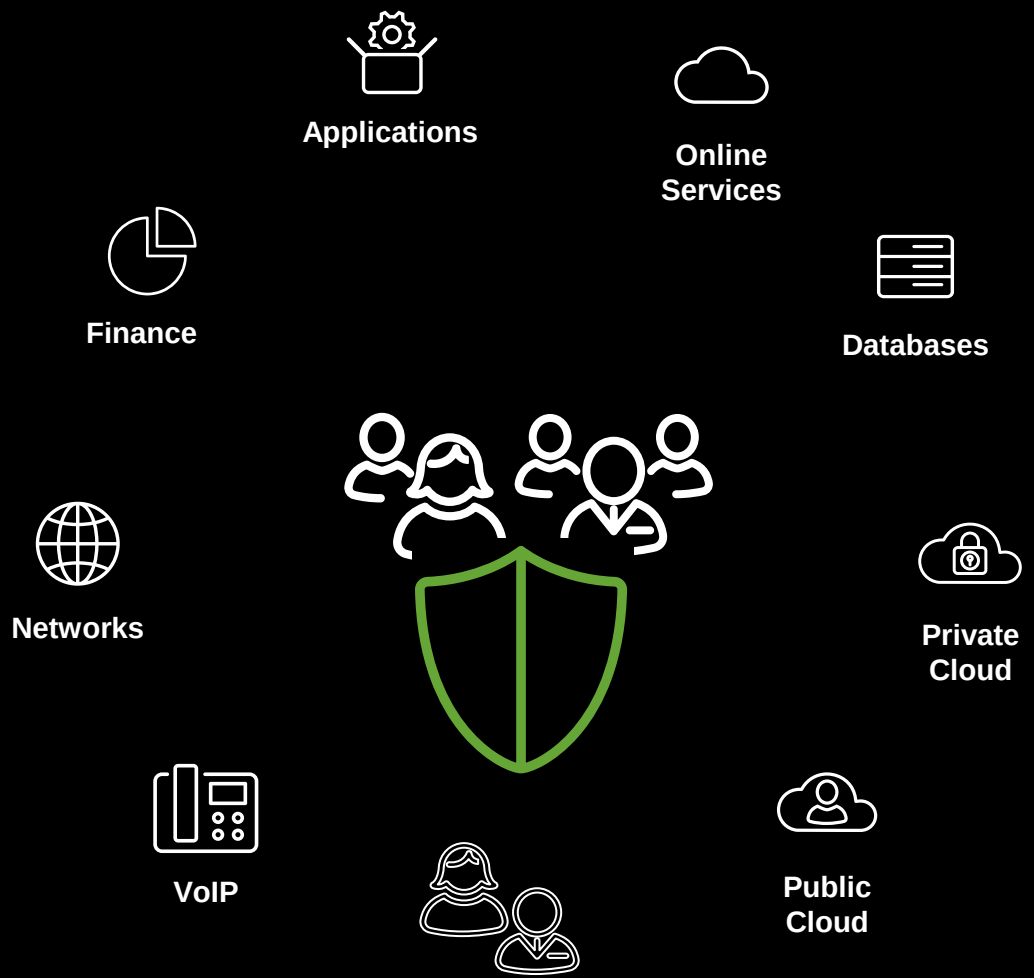


splunk> .conf18

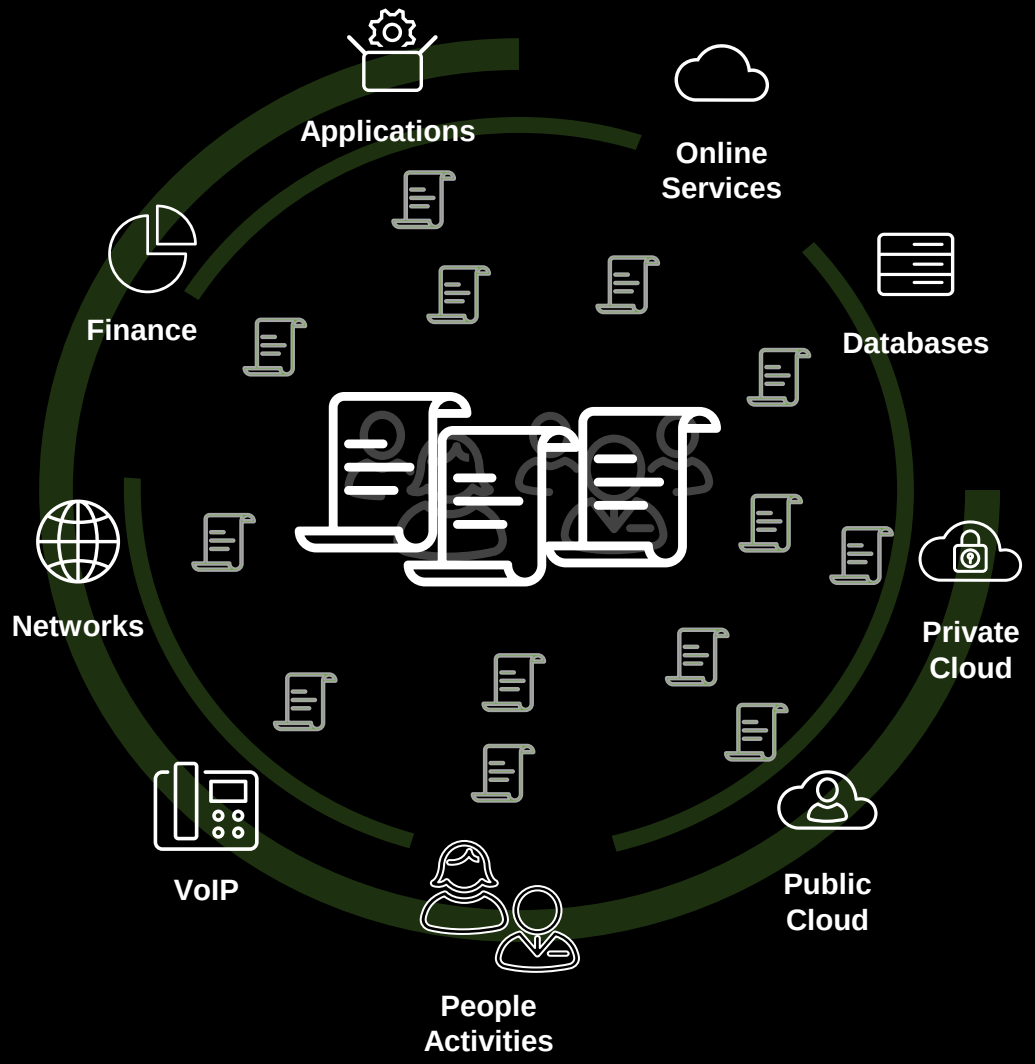
# The Problem:

**Second guessing your own company**

# Security = Knowing Your Company Well

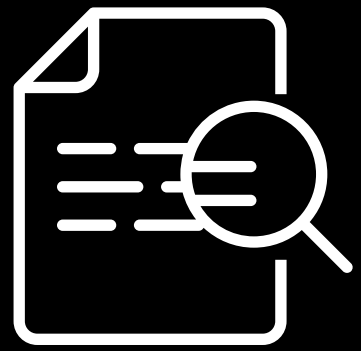


# Having Data ≠ Security



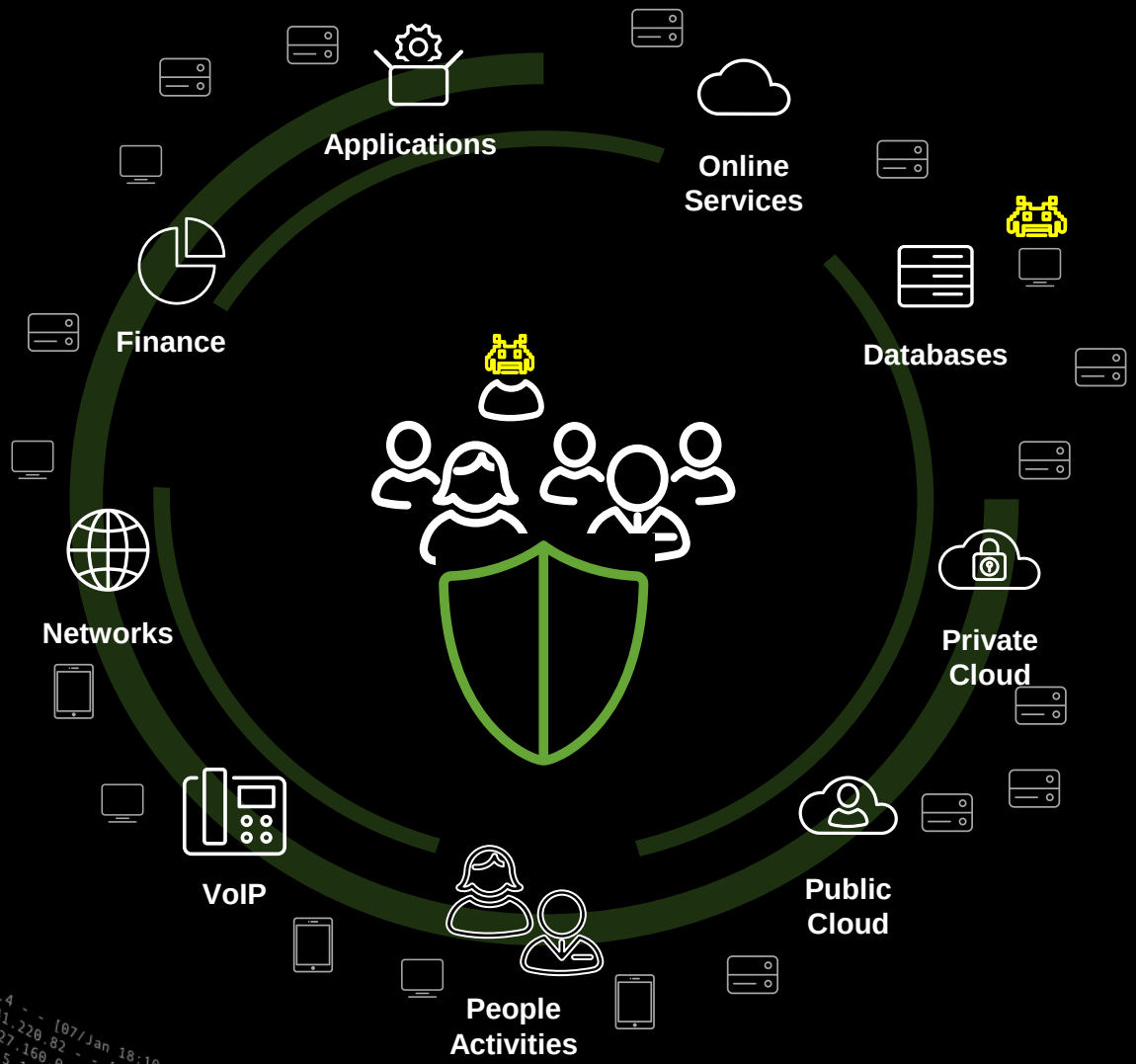
# Organized Data → Security



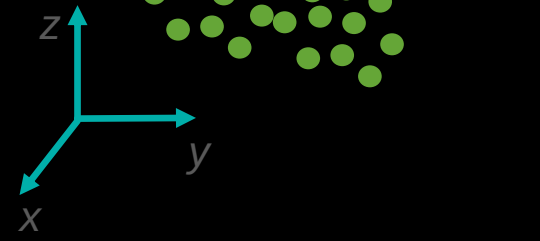


# THE Security Question

## Which entities act oddly?

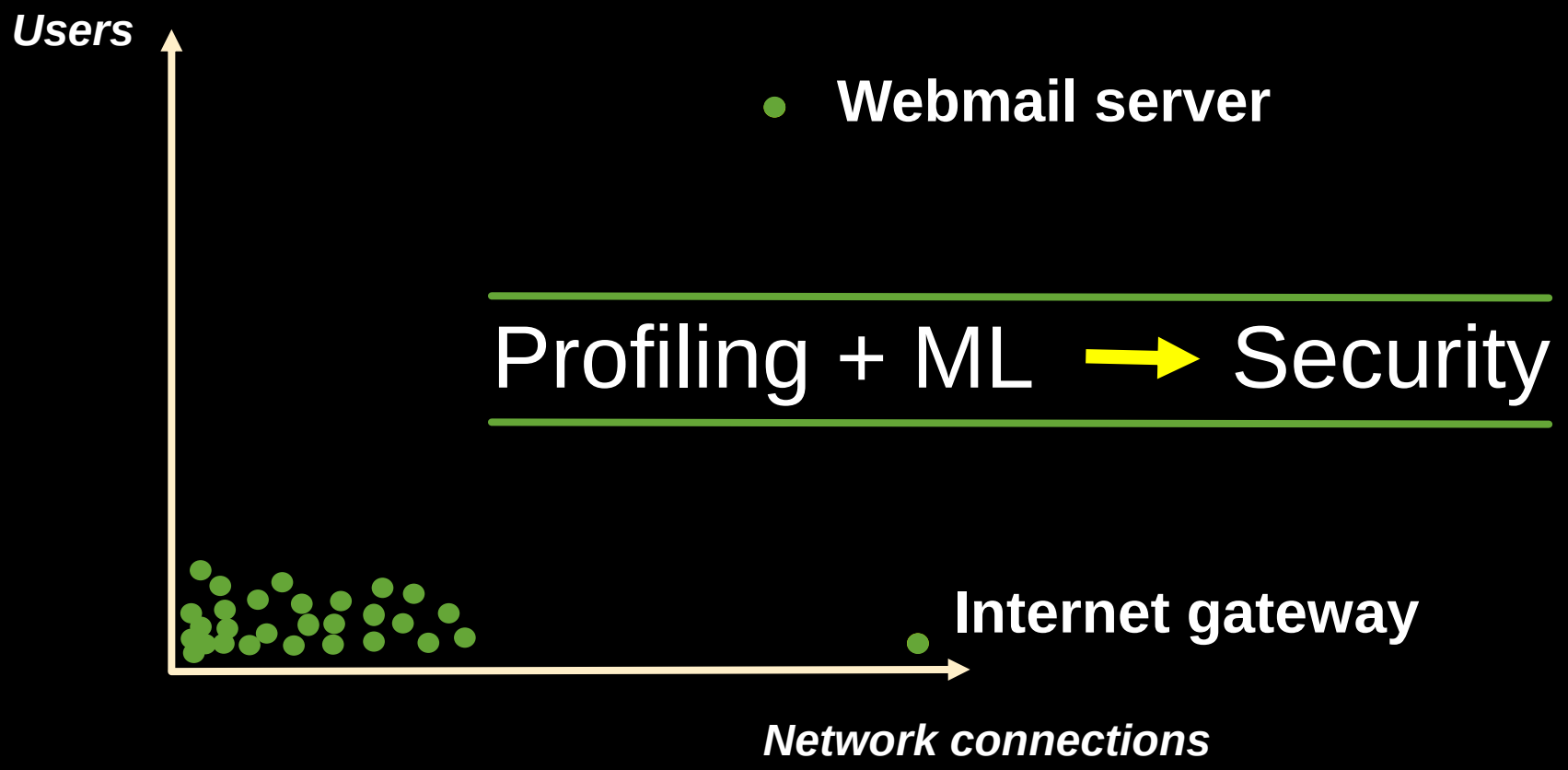



splunk> UBA



# ML ≠ Security

## Illustration: Device Clustering




# Obstacles:

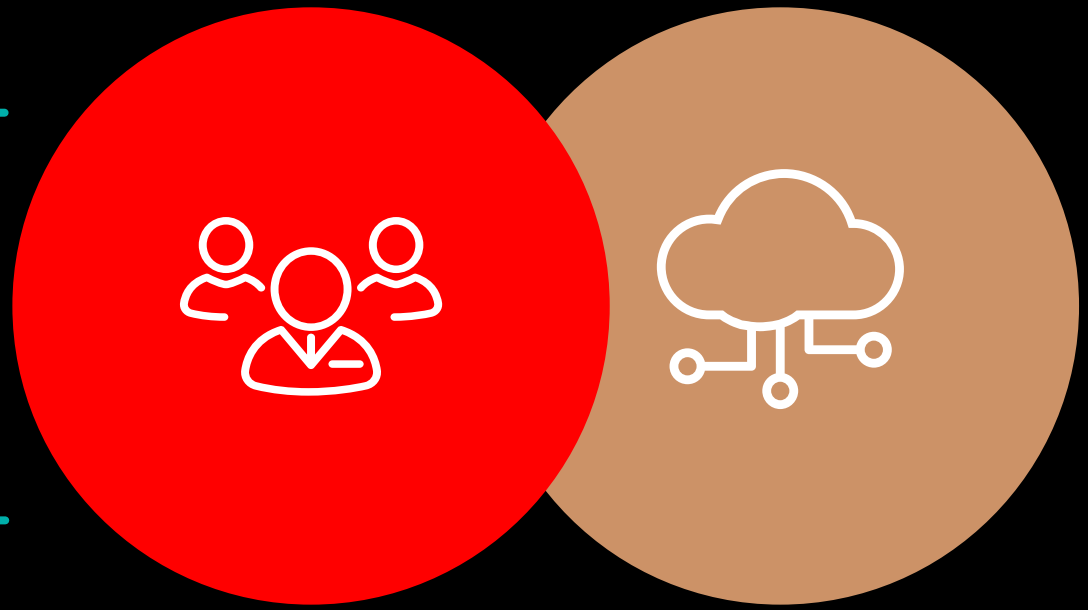
## Profiling challenges

# Profiling – The Missing Link

Undocumented

**Human Involvement**

Crowd sourced



No perimeters

**Technology Evolution**

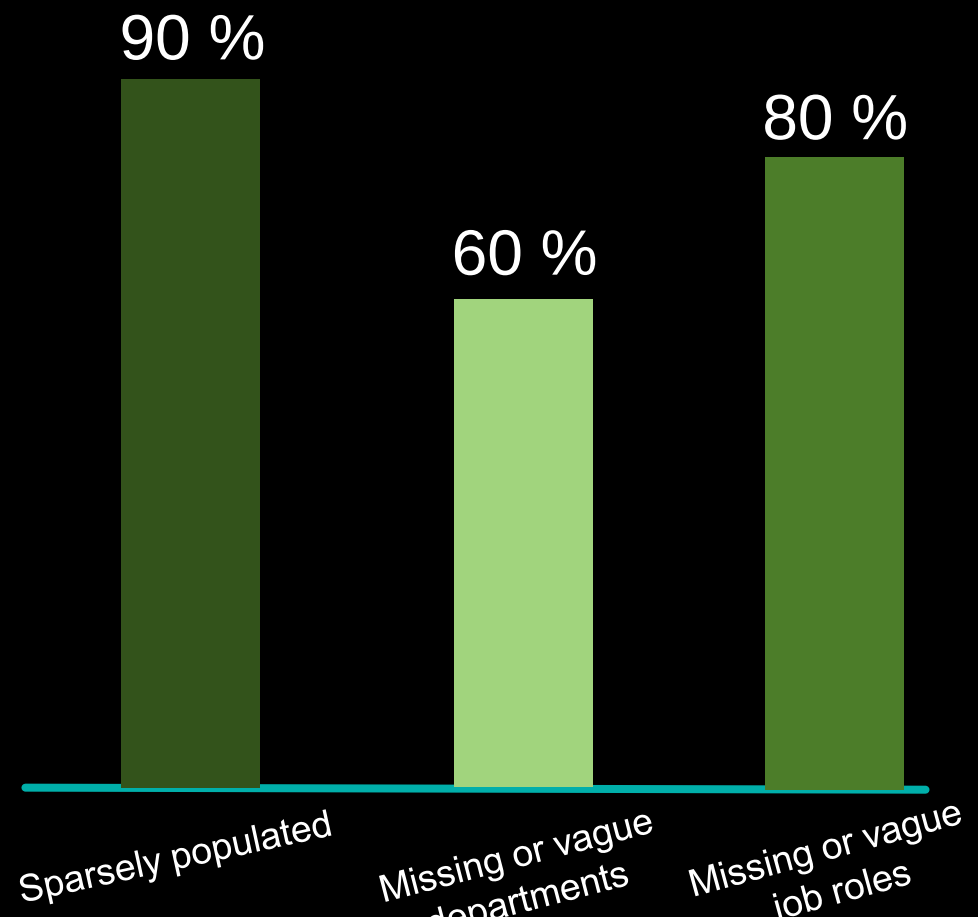
Integration

# Profiling – The Missing Link

## Human Factor

| User OU names | Num. users |
|---|---|
| NORMAL | 6031 |
| WORKSTATION | 52 |
| ..... | |
| TERM EIF 12*** | 1 |
| EIF CHI 2***** | 1 |
| EIF: H**** | 1 |
| Nonprofit | 1 |
| EIF: ***** | 1 |
| Left December *** | 1 |
| NY Term*** | 1 |
| EIF **** | 1 |
| Dean | 1 |
| ***176 FR-PAR | 1 |
| Term Form NY **** | 1 |
| ..... | |

Sample: ”Human touch” in LDAP Data




LDAP Data

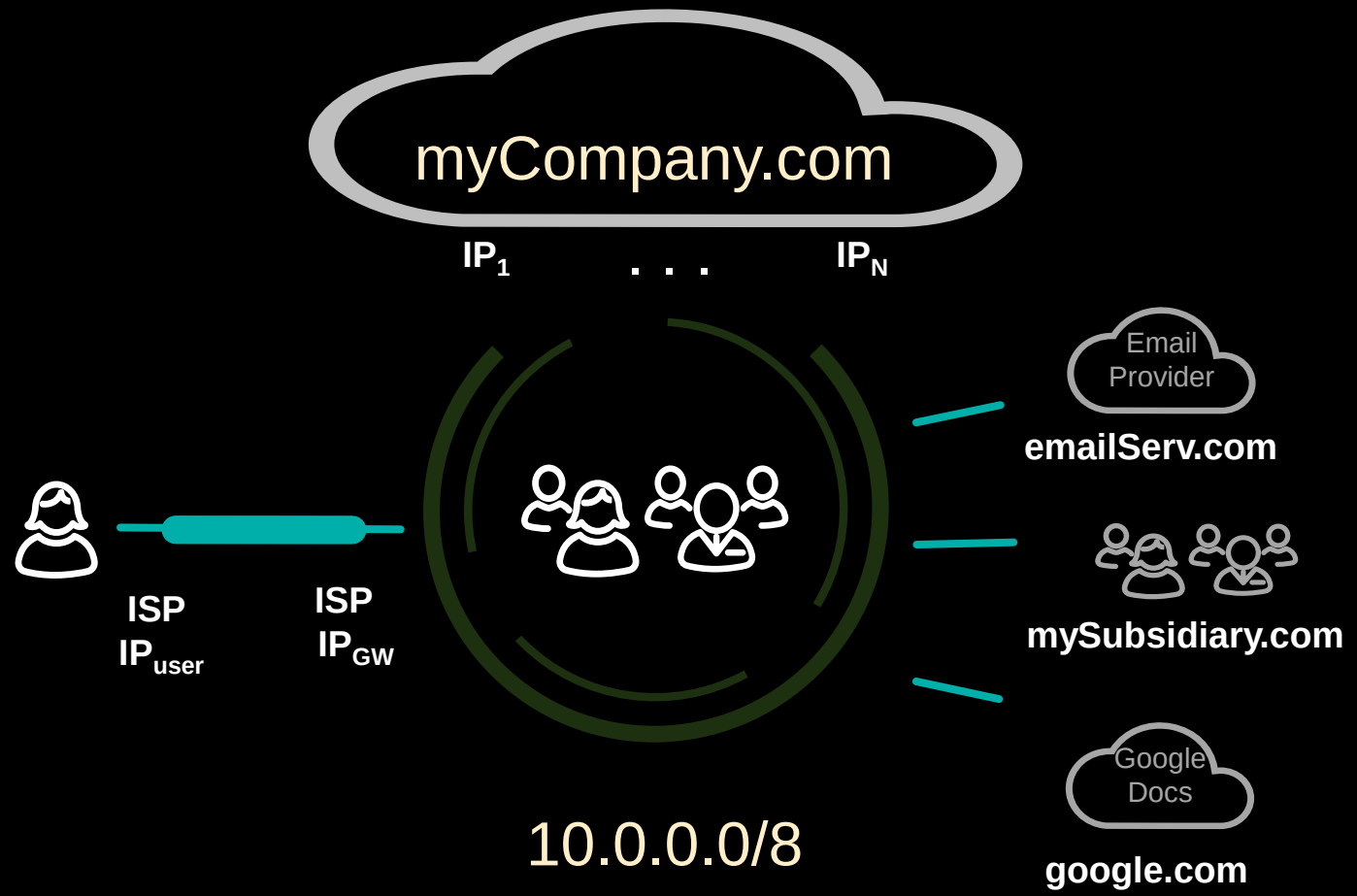
# Profiling – The Missing Link

## Technology Aspects

## “Impossible” questions:

- What are your domains?
- What are your IP ranges?

- What are all your servers?
- Who are the critical people?
- What are the critical docs?



# Profiling in Splunk UBA

**Enrichment of security context via algorithms**

splunk> .conf18

# Profiling Overview

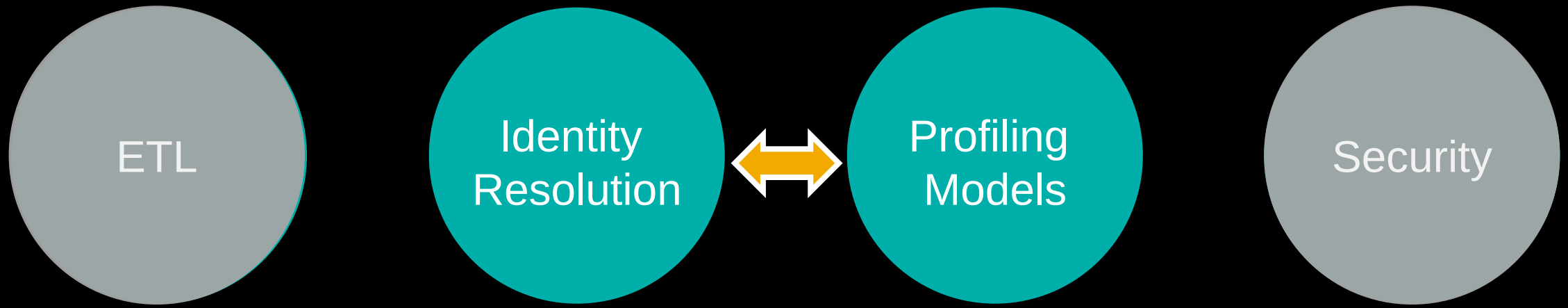
**10.59.42.7**

- WiFi address
- Guest network
- …

- Person
- Attack capable
- Owns deviceX
- …

- Server
- Email Server
- Web accessible
- …

# Architecture



# Architecture


**Entities**

multiline
cef
Active Directory
Office365
Infoblox
DNS
DHCP
snare
evtx
Windows
Syslog
NetScaler
Weblog
AWS
Cisco
FireEye
DropBox
Box
Juniper
SonicWall
SNORT
Palo Alto Networks
Auth
Avro
Symantec
BRO
Appcelerator
Imperva
BlueCoat
Okta
McAfee
ForeScout
Checkpoint

… plus Splunk TAs

# Architecture

Identity Resolution

## Augmentation

WileE

RRlaptop

10.251.0.24

AD

Nov 10 06:00:18 SCL-DC01.acme.com/10.115.16.5/1.9.130.1 MSWinEventLog,1,Security,896530380,Sat Nov 10 06:00:17 2018,4769,Microsoft-Windows-Security-Auditing, ACME\RRlaptop$@ACME.COM,N/A,Success Audit,SCL-DC01.acme.com,Kerberos Service Ticket Operations,,A Kerberos service ticket was requested. Account Information: Account Name: RRlaptop$@ACME.COM Account Domain: ACME Logon GUID: {7DBD8A74-923E-5F47-C3D1-B70217F23443} Service Name: S-1-5-21-196153179-1972187586-475923621-502 Network Information: Client Address: ::ffff:10.251.0.24 Client Port: 62929 Additional Information: Ticket Options: 0x60810010 Ticket Encryption: 0x17 Failure Code: 0x0 Transited Services: -

RRlaptop$@acme.com

::ffff:10.251.0.24

AD

Nov 10 06:03:19 SCL-DC01.acme.com/10.115.16.5/1.9.130.1 MSWinEventLog,1,Security,896622469,Sat Nov 10 06:03:18 2018,4624,Microsoft-Windows-Security-Auditing,ACME\WileE,N/A,Success Audit,SCL-DC01.acme.com,Logon,,An account was successfully logged on. Subject: Security ID: S-1-0-0 Account Name: - Account Domain: - Logon ID: 0x0 Logon Type: 3 New Logon: Security ID: S-1-5-21-196452179-1272397586-475923621-7718 Account Name: WileE Account Domain: ACME, Logon ID: 0xb068c7e29 Logon GUID: {068DDCB8-F1FB-BA08- Process ID: 0x0 Process Name: - Network Information: Workstation Name: Source Network Address: 10.251.0.24 Source Port: 56023

WileE

10.251.0.24

\+ RRlaptop

PAN

1541829805 pansplunk3 indexer_guid=6527EB1D-75F2-47E7-8196- fdf0fa5371f7788075470964c3cb9c9a5@@sourcetype=pan_threat@@datamodel=Intrusion_Detection c-delivery-networks@@dest=208.111.179.93(https-208-111-179- .93(https-208-111-179-93.sea.llnw.net)@@dest_dns=208.111.179.93(https-208-111-179-93.sea.llnw.net)@@dest_port=443@@dv odsplunk3@@ids_type=network@@product=Firewall@@rule=Standard HTTP-HTTPS@@severity=informational src_ip= 10.251.0.24(unresolved)@@src_dns= 10.251.0.24(unresolved)@@src_port=52390 r=PaloAlto@@tag=attack@@tag=ids@@tag=network

linkedin-base

10.251.0.24

system

\+ RRlaptop + WileE

splunk> .conf18

# Architecture

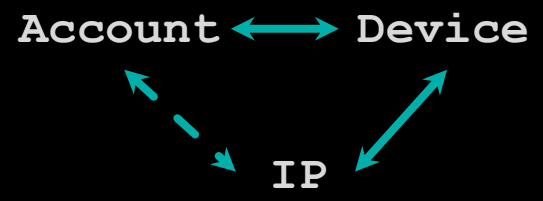
Profiling Models

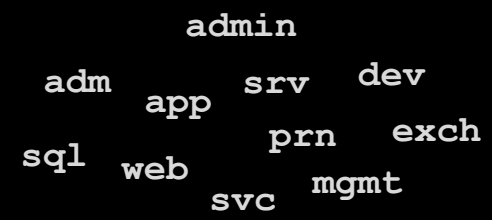
**Context**

FEATURES

| | |
|---|---|
| Domain Expertise | • Authentication types<br>• Credentials<br>• Processes<br>• Special events … |
| Associations | • Assoc. set diversity<br>• Duration<br>• Persistence … |
| Naming Analysis | • Industry practices<br>• Customer insights … |

.conf ’17 talks




Successful discovery of entire *critical infrastructures*.

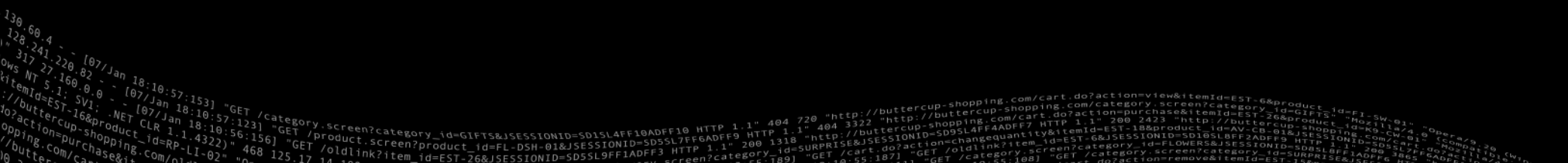
splunk> .conf18

# Efficacy

## Example: Device Profiling

| | |
|---|---|
| Total devices | 13,998 |
| Profiled devices | 7135 |

| Dominant features | Num of devices : |
|---|---|
| Domain expertise | 352 |
| Device – IP associations | 6287 |
| Device – User associations | 742 |
| Naming analysis | 1094 |



# Example: Association Features (Simplified)

- Idea

***Similar*** features, ***diverse*** profile discoveries

- Features
  - `Observation duration`
  - `Daily average of associated entities`
  - `Daily max of associated entities`

- Profiles to be discovered
  - `NAT IP addresses`
  - `Multihomed devices`

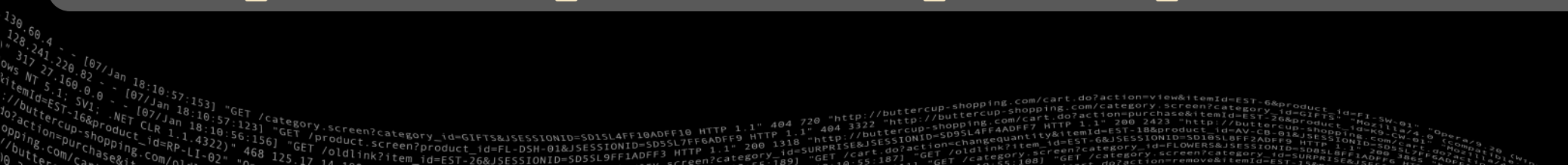
splunk> .conf18

# Example: Association Features (Simplified)

## Discovery of NAT addresses

```
index="yourAD" EventCode=4624 OR EventCode=4769 |
rex field=_raw "New Logon:\s+Security ID:\s+(?<sid>[^\r\n]+)\s+Account Name:\s+(?<new_logon_account_name>[^\r\n]+)\s+Account
Domain:\s+(?<new_logon_domain>[^\r\n]+)\s+Logon ID" |
rex field=_raw "Logon Process:\s+(?<logon_process>[^\r\n]+)\s+" |
rex field=_raw "Workstation Name:\s+(?<workstation_name>[^\r\n]+)\s+" |
rex field=_raw "Source Network Address:\s+(::ffff:)?(?<source_address>\S+)\s+Source" |
rex field=_raw "Service Name:\s+(?<service_name>[^\r\n]+)\s+Service ID" |
rex field=_raw "Account Name:\s+(?<account_name>[^@\r\n]+)\S*\s+\S+" |
rex field=_raw "Client Address:\s+(::ffff:)?(?<client_address>\S+)\s+Client" |
search (EventCode=4624
              AND ((logon_process=Kerberos AND new_logon_account_name="*$") OR (logon_process=NtlmSsp AND workstation_name != ""))
              AND source_address!="-" AND source_address != "127.0.0.1" AND source_address!="::1")
        OR
        (EventCode=4769
            AND service_name=krbtgt AND account_name = "*$*"
            AND client_address != "-" AND client_address != "127.0.0.1" AND client_address != "::1") |
eval device_name_4624 = if(logon_process == "NtLmSsp", workstation_name,  mvindex(split(new_logon_account_name, "$"), 0)) |
eval device_name = if(EventCode=4624, lower(device_name_4624), lower(mvindex(split(account_name, "$"), 0))) |
eval address = if(EventCode=4624, source_address, client_address) |eval date_hm = strftime(_time, "%m/%d/%Y") |
stats dc(device_name) as numdev by address date_hm |
```

```
stats  dc(date_hm) as num_days, avg(numdev) as avg_dev, max(numdev) as max_dev
by address |
where avg_dev > 1 AND avg_dev >= 0.8 * total_dev AND num_days > 2
```



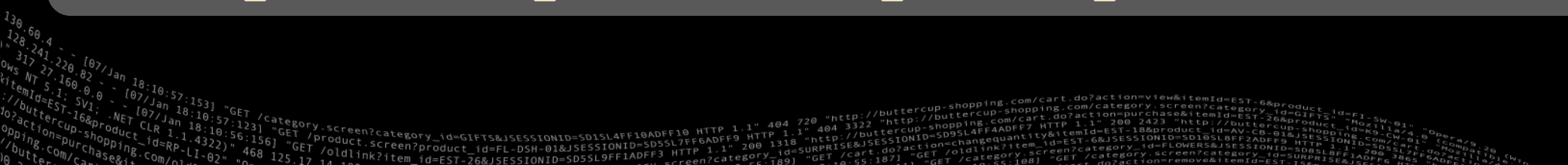
splunk> .conf18

# Example: Association Features (Simplified)

## Discovery of multihomed servers

```
index="yourAD" EventCode=4624 OR EventCode=4769 |
rex field=_raw "New Logon:\s+Security ID:\s+(?<sid>[^\r\n]+)\s+Account Name:\s+(?<new_logon_account_name>[^\r\n]+)\s+Account
Domain:\s+(?<new_logon_domain>[^\r\n]+)\s+Logon ID" |
rex field=_raw "Logon Process:\s+(?<logon_process>[^\r\n]+)\s+" |
rex field=_raw "Workstation Name:\s+(?<workstation_name>[^\r\n]+)\s+" |
rex field=_raw "Source Network Address:\s+(::ffff:)?(?<source_address>\S+)\s+Source" |
rex field=_raw "Service Name:\s+(?<service_name>[^\r\n]+)\s+Service ID" |
rex field=_raw "Account Name:\s+(?<account_name>[^@\r\n]+)\S*\s+\S+" |
rex field=_raw "Client Address:\s+(::ffff:)?(?<client_address>\S+)\s+Client" |
search (EventCode=4624
              AND ((logon_process=Kerberos AND new_logon_account_name="*$") OR (logon_process=NtlmSsp AND workstation_name != ""))
              AND source_address!="-" AND source_address != "127.0.0.1" AND source_address!="::1")
        OR
        (EventCode=4769
            AND service_name=krbtgt AND account_name = "*$*"
            AND client_address != "-" AND client_address != "127.0.0.1" AND client_address != "::1") |
eval device_name_4624 = if(logon_process == "NtLmSsp", workstation_name,  mvindex(split(new_logon_account_name, "$"), 0)) |
eval device_name = if(EventCode=4624, lower(device_name_4624), lower(mvindex(split(account_name, "$"), 0))) |
eval address = if(EventCode=4624, source_address, client_address) |eval date_hm = strftime(_time, "%m/%d/%Y") |
stats dc(address) as numaddr by device_name date_hm |
```

```
stats dc(date_hm) as num_days, avg(numaddr) as avg_ip, max(numaddr) as max_ip
by device_name |
where avg_ip > 1 AND avg_ip >= 0.7 * total_ip AND num_days > 2
```



splunk> .conf18

# Results

## Simplified profiling by association features

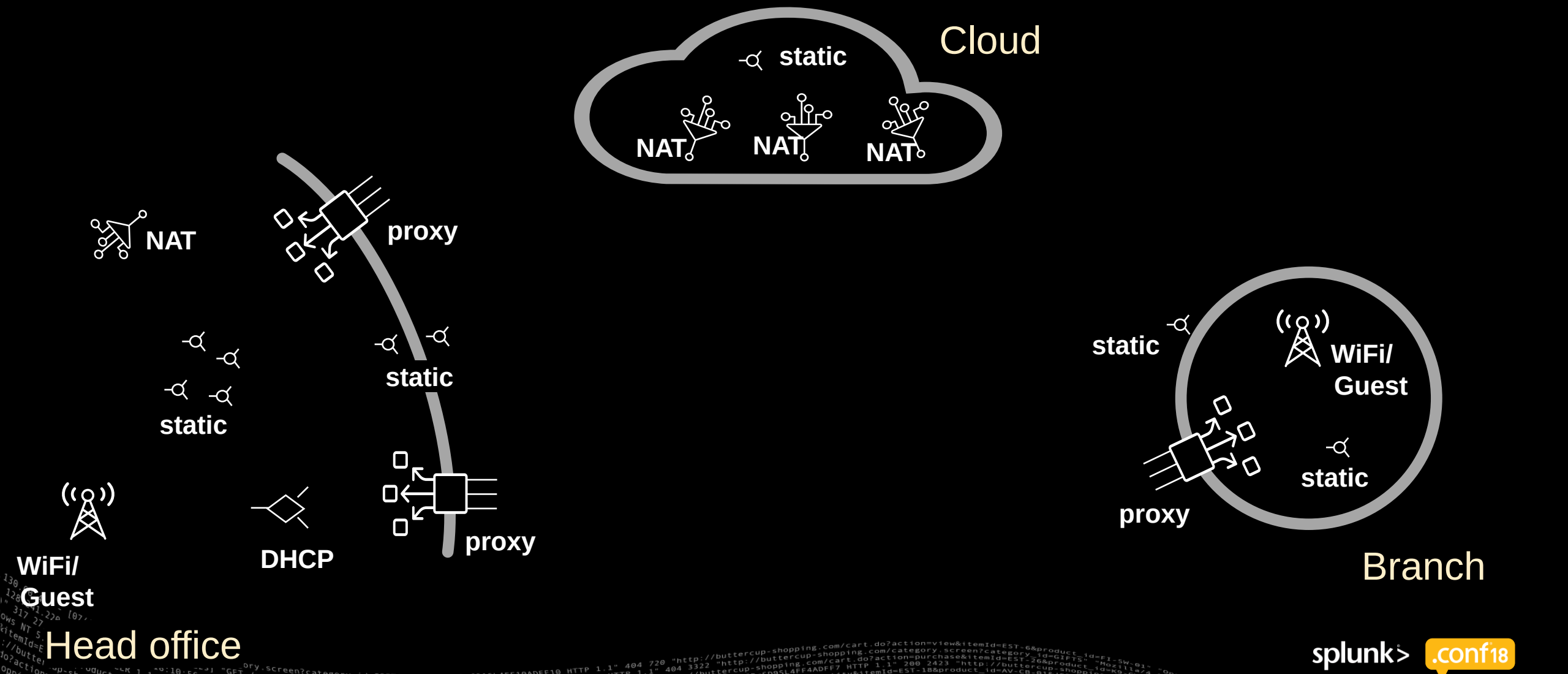
NAT Addresses

| address | num_days | avg_dev | max_dev |
|---|---|---|---|
| 10.[redacted].30 | 11 | 108 | 129 |
| 10.[redacted].31 | 11 | 105.63636363636364 | 126 |
| 10.[redacted]20 | 11 | 72.0909090909091 | 89 |
| 10.[redacted]18 | 12 | 15.75 | 17 |
| 168.[redacted].211 | 12 | 10.75 | 13 |
| 168.[redacted].212 | 12 | 9.166666666666666 | 10 |
| 168.[redacted].210 | 12 | 5 | 5 |

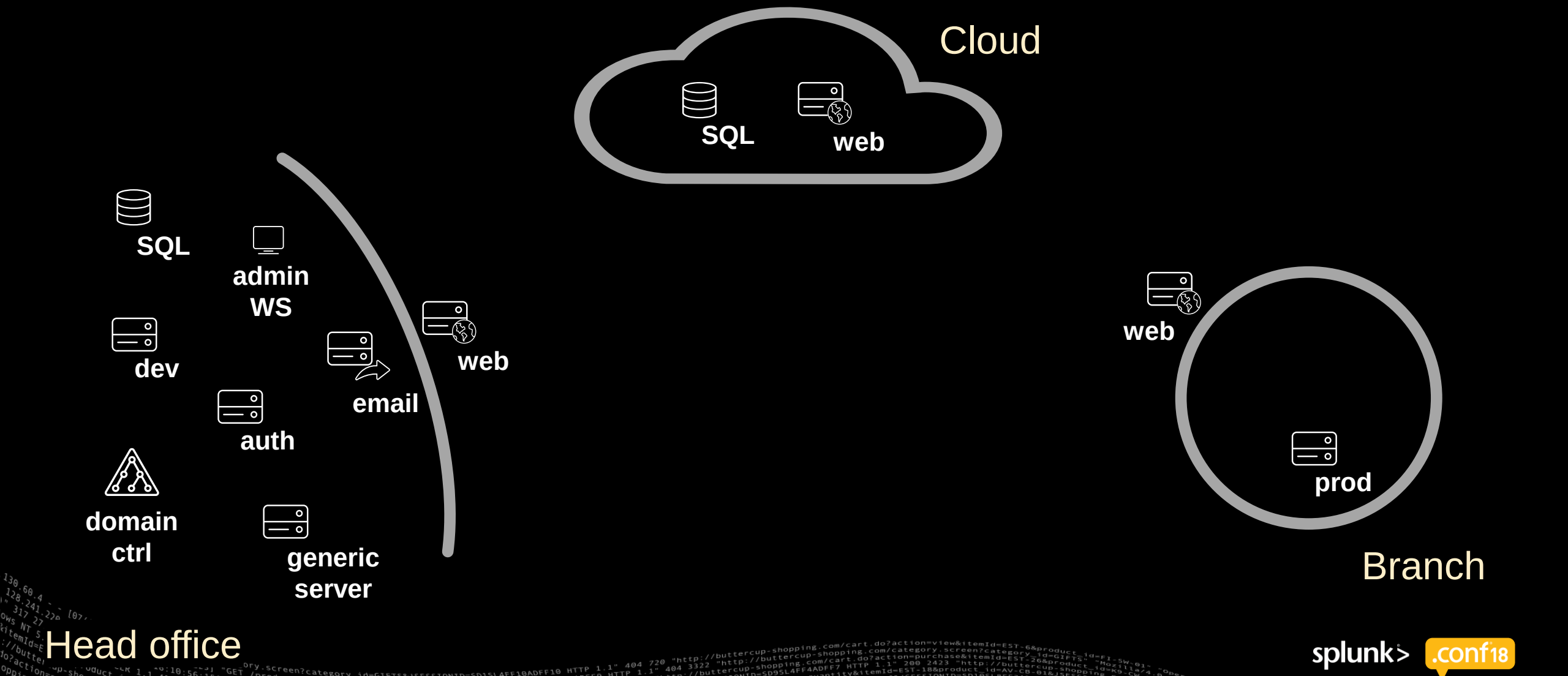
Multihomed Servers

| device_name | num_days | avg_ip | max_ip |
|---|---|---|---|
| cnd[redacted] | 6 | 4.5 | 6 |
| cnd[redacted] | 5 | 4.2 | 6 |
| cnu[redacted] | 5 | 3.6 | 5 |
| cnu[redacted] | 5 | 3.6 | 5 |
| cnu[redacted] | 7 | 3.5714285714285716 | 5 |
| cnu[redacted] | 3 | 3.3333333333333335 | 4 |

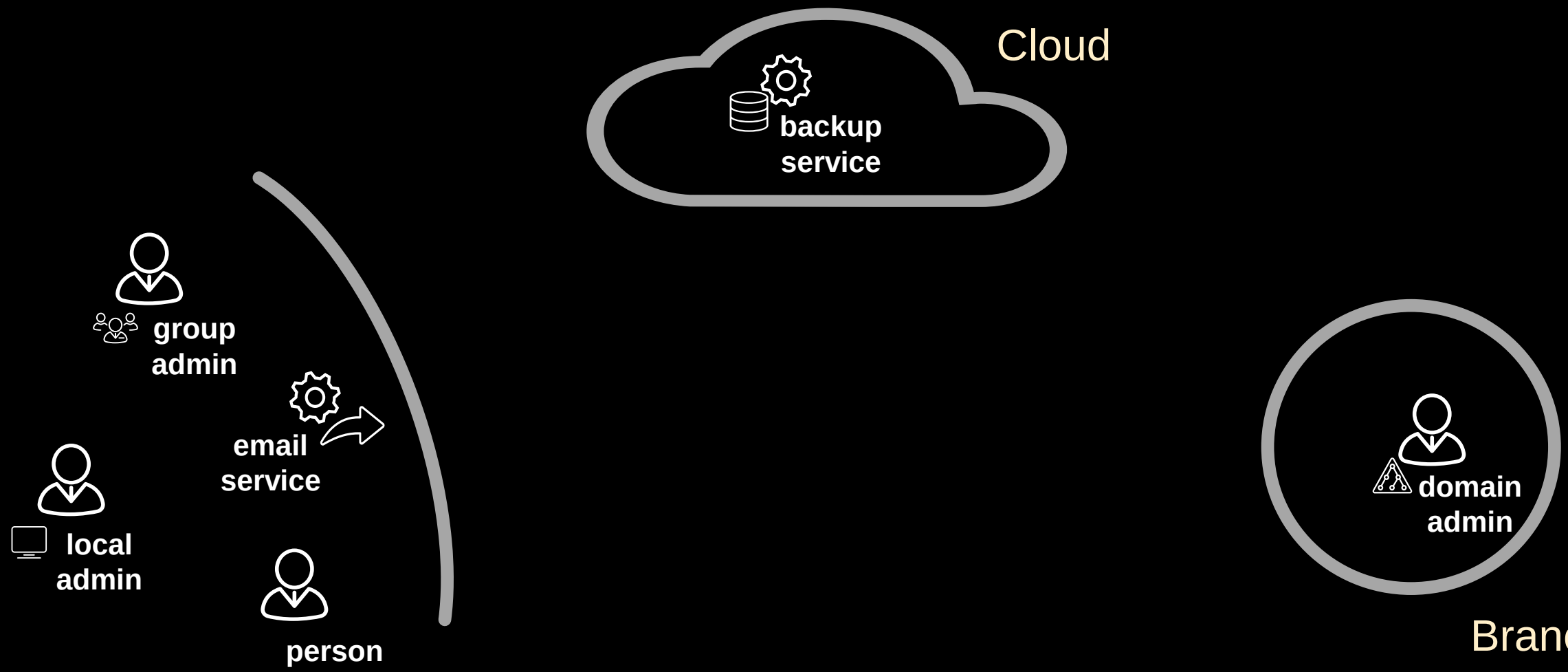
# UBA Profiles



# UBA Profiles



# UBA Profiles … so far



# UBA Profiles

## IP Addresses

**IP "Inadequate" For IR**

**Proxy IP address**
- Public
- Private

**NAT IP address**
- Public
- Private

**High-rate DHCP IP address (guest network)**
- Public
- Private

**Static IP address**
- Public
- Private

## Domains (AD & DNS)

AD Domains IOC table

**Company's AD domains**

**Company's DNS domains**

## Devices

Device Profiles table

**Domain Controller**

**Web server**
**Email servers**
**DNS server**
**DHCP server**
**DMZ server**
**Auth server**

**Backup server**
**SQL servers**
**Development server**
**Print server**
**Production server**
**….**

**Static IP server**
**Public IP server**
**Private IP server**

**Multi-homed server**

**Admin's workstation**

## Accounts/Users

User Profiles table

~~**Admin account**~~
~~**Domain Admin**~~
~~**Local (device) admin**~~
~~**Generic admin in AD Domain**~~
**Attack-capable account**

**Service account**
**Email Service account**
**Security Service account**
**SQL Service account**
**Web Service account**
**…**
**Batch account**

**Web-Login account**

**Interactive-Login account (person)**

splunk> .conf18

# Key Takeaways

**UBA profiling**

1. Profiling: Turns data to clear context
   - Who are your people, devices, infrastructure?
   - Automated by machine learning
   - Data driven and accurate
2. Easy to build security use cases
   - For human analysts
   - For other algorithms
3. Can humans do it well?  No!

# Have a question?

1. Stanislav Miskovic
   - Email: smiskovic@splunk.com
2. Prasoon Shukla
   - Email: pshukla@splunk.com
3. Dimitrios Terzis
   - Email: dterzis@splunk.com

# Thank You

**Don't forget to rate this session in the .conf18 mobile app**