
jtsec
BEYOND IT SECURITY

CC Web Scraper 2.0

Index

- Recap on CC Web Scrapper v1.0
- What's new in CC Web Scrapper v2.0
- Implementation and difficulties while parsing
- Obtaining and correlating archived data from CBs
- Technical conclusions
- Statistics
- Final conclusions

Index

- **Recap on CC Web Scrapper v1.0**
- What's new in CC Web Scrapper v2.0
- Implementation and difficulties while parsing
- Obtaining and correlating archived data from CBs
- Technical conclusions
- Statistics
- Final conclusions

Recap on version 1.0

- The CC Web Scraper was first presented on ICCC 2018.
- A month later, the first report generated by the tool was released.
  - The second report was released on February, 2019 and the third on last September.

- https://www.jtsec.es/blog-entry/21/common-criteria-statistics-report-for-q1-q2-q3-of-2018
- https://www.jtsec.es/blog-entry/25/common-criteria-statistics-report-for-2018
- 2019 Q3 report has been just released!

Index

- Recap on CC Web Scrapper v1.0
- **What's new in CC Web Scrapper v2.0**
- Implementation and difficulties while parsing
- Obtaining and correlating archived data from CBs
- Technical conclusions
- Statistics
- Final conclusions

What's new?

- The main downside of the scraper that was discussed last year was its inability to parse the products directly from Certification Bodies.

What's new?

- The scraper now gets info from every Certification Body!
- And it is able to compare the results with those from Common Criteria Portal.

Index

- Recap on CC Web Scrapper v1.0
- What's new in CC Web Scrapper v2.0
- **Implementation and difficulties while parsing**
- Obtaining and correlating archived data from CBs
- Technical conclusions
- Statistics
- Final conclusions

Implementation

- To parse the Certification Bodies websites, we have included a new manager within the tool that individually collects information for each CB.

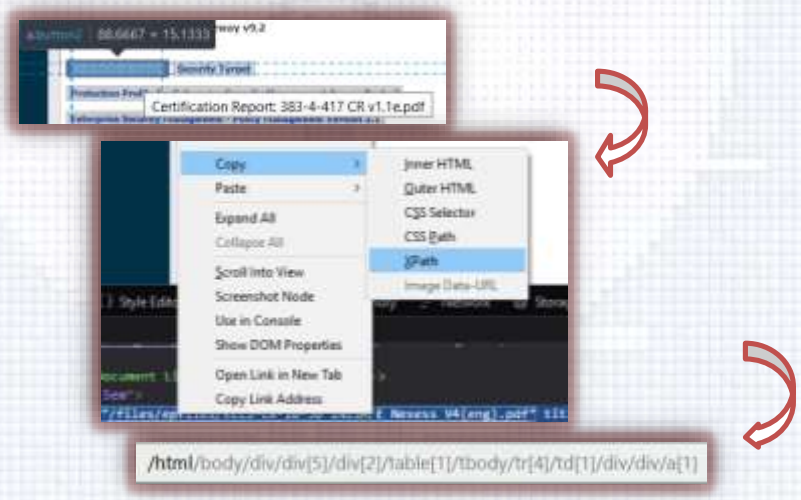
```
                    CC Portal Web Scraper
######################################

## Running Certification Bodies scraper
        ## Running ACSC scraper (Austra
        ## Running ANSSI scraper (Franc
        ## Running BSI scraper (Germany
        ## Running STQC scraper (India)
        ## Running OCSI scraper (Italy)
        ## Running IPA scraper (Japan)
        ## Running MyCC scraper (Malays
        ## Running NSCIB scraper (Nethe
        ## Running SERTIT scraper (Norw
        ## Running ITSCC scraper (Reput
        ## Running CSA scraper (Singapo
        ## Running CCN scraper (Spain)
        ## Running CSEC scraper (Sweden
        ## Running NIAP scraper (United
```

```
36      cc_products.append(acsc_scraper.get_acsc_elements())        # Australia / New Zeland
37                                                                  # Canada (CC Portal)
38      cc_products.append(anssi_scraper.get_anssi_elements())      # France
39      cc_products.append(bsi_scraper.get_bsi_elements())          # Germany
40      cc_products.append(stqc_scraper.get_stqc_elements())        # India
41      cc_products.append(ocsi_scraper.get_ocsi_elements())        # Italy
42      cc_products.append(ipa_scraper.get_ipa_elements())          # Japan
43      cc_products.append(mycc_scraper.get_mycc_elements())        # Malaysia
44      cc_products.append(nscib_scraper.get_nscib_elements())      # Netherlands
45      cc_products.append(sertit_scraper.get_sertit_elements())    # Norway
46      cc_products.append(itscc_scraper.get_itscc_elements())      # Rep. of Korea
47      cc_products.append(csa_scraper.get_csa_elements())          # Singapore
48      cc_products.append(ccn_scraper.get_ccn_elements())          # Spain
49      cc_products.append(csec_scraper.get_csec_elements())        # Sweden
50                                                                  # Turkey (CC Portal)
51                                                                  # United Kingdom (CC Portal)
52      cc_products.append(niap_scraper.get_niap_elements())        # United States
```

Implementation

- First version of the scraper was based on xpath.
- Although this approach worked when parsing the CC Portal, it fails on most of the Certification Bodies websites.



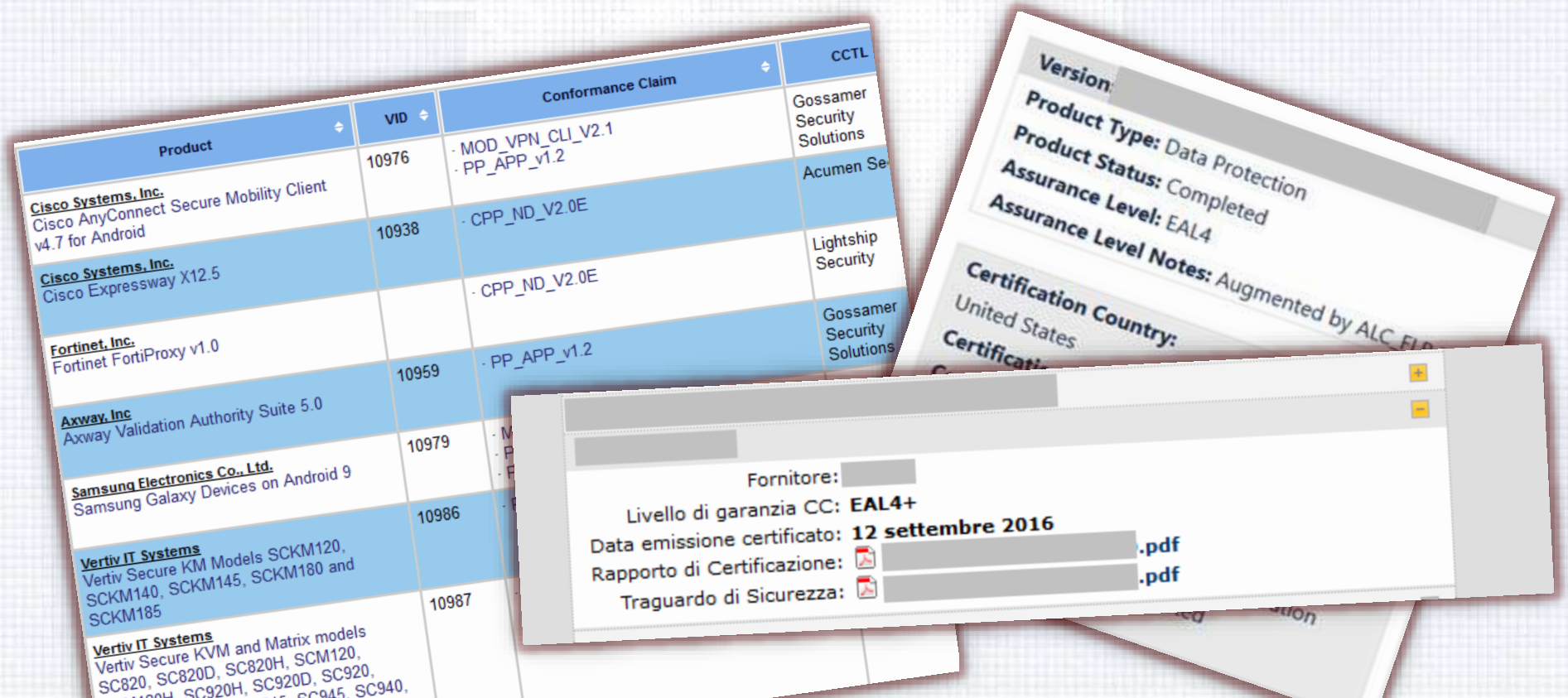
- BeautifulSoup performs better

CBs that rely on CC Portal

- The UK does not have a list of products
  - They publish certifications through CC Portal.
- Canada does the same, but presents a list with some data.
- Turkey has a PDF list but all links lead to CC portal.
  - The only one that doesn't is a broken link.

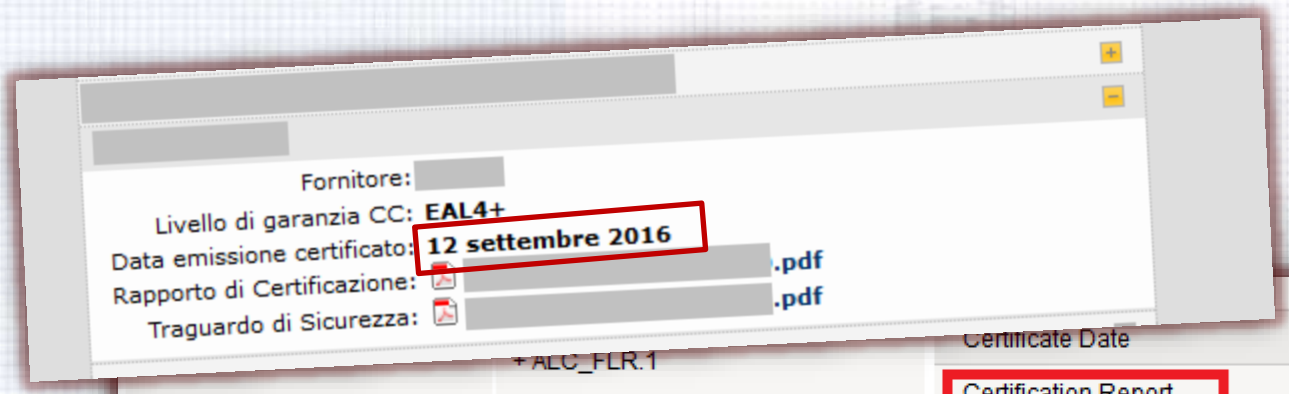
Difficulties while parsing

- Some Certification Bodies have simple well structured websites, which are easy to parse with the new functionality.



Difficulties while parsing

- Worst case scenario, there are some minor differences such as different date formats or different names for the same thing.



+ ALC_FLR.1	
Certificate Date	2015-06-17
Certificate Report	Certification Report (pdf)
Certificate	Certificate (pdf) Certificate (pdf)

Certificate Date	2019-08-15
Certification Report	Certification Report (pdf)
Certificate	Certificate CCRA (pdf) Certificate SOGIS (pdf) Certificate EA MLA (pdf)

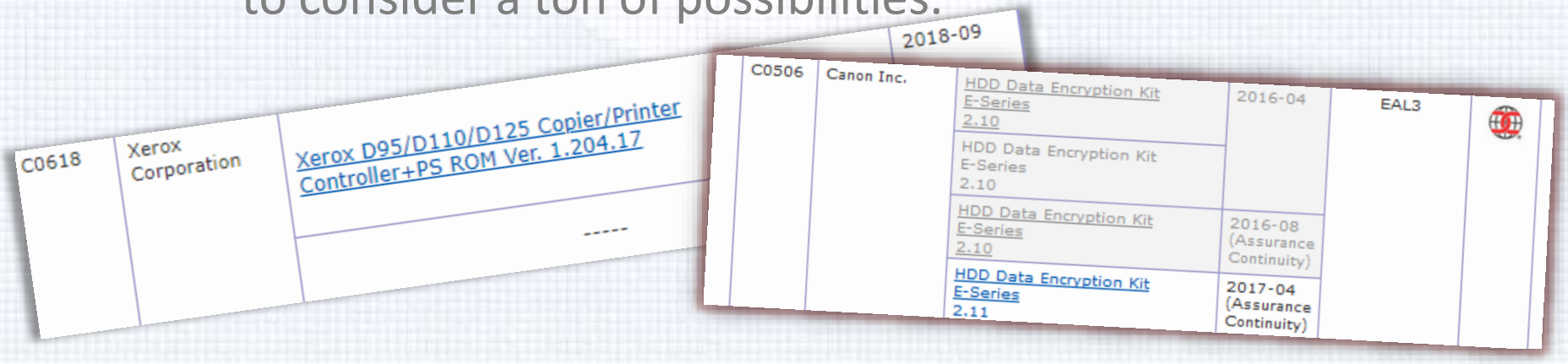
Difficulties while parsing

- Another problem was presented while parsing ANSSI's website.
- After several continued requests, the server starts denying connections
- We placed a timeout and problem solved…

Anti - DoS

Difficulties while parsing

- With IPA the problem took some more effort.
- They maintain archived certification documents in the same table but with a different format.
- Some products have several versions, names, etc. making it unpredictable when parsing and forcing to consider a ton of possibilities.



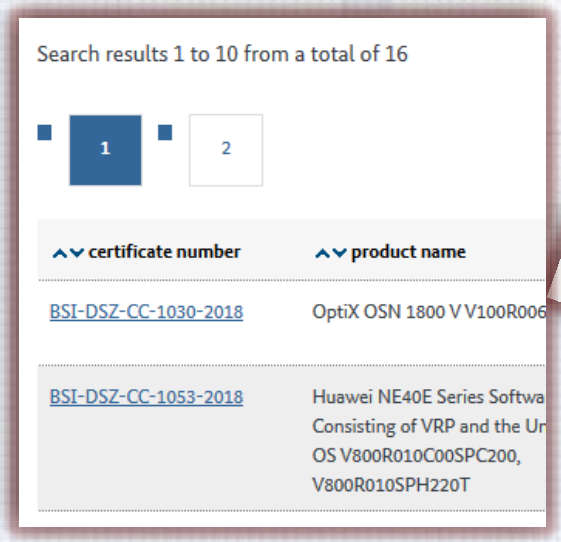
C0618	Xerox Corporation	Xerox D95/D110/D125 Copier/Printer Controller+PS ROM Ver. 1.204.17	2018-09		

C0506	Canon Inc.	HDD Data Encryption Kit E-Series 2.10	2016-04	EAL3	
		HDD Data Encryption Kit E-Series 2.10			
		HDD Data Encryption Kit E-Series 2.10	2016-08 (Assurance Continuity)		
		HDD Data Encryption Kit E-Series 2.11	2017-04 (Assurance Continuity)		

Difficulties while parsing

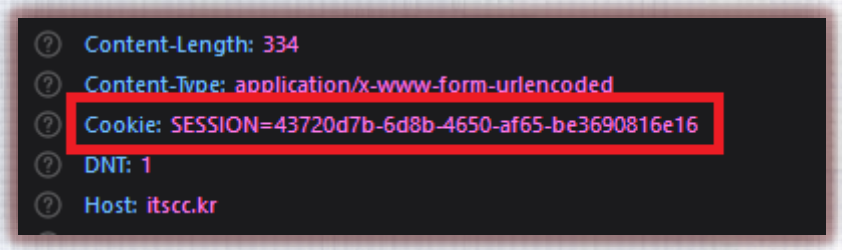
- The difficulties grew larger as more complex websites were being scraped.

- The BSI divides its products in 10 categories
  - Which are then divided in several subcategories…
  - Which are divided by pages when they are large…
  - Or empty when they have no products…



Difficulties while parsing

- Korean Certification Body has several security controls on its website which had to be tackled to parse their website.



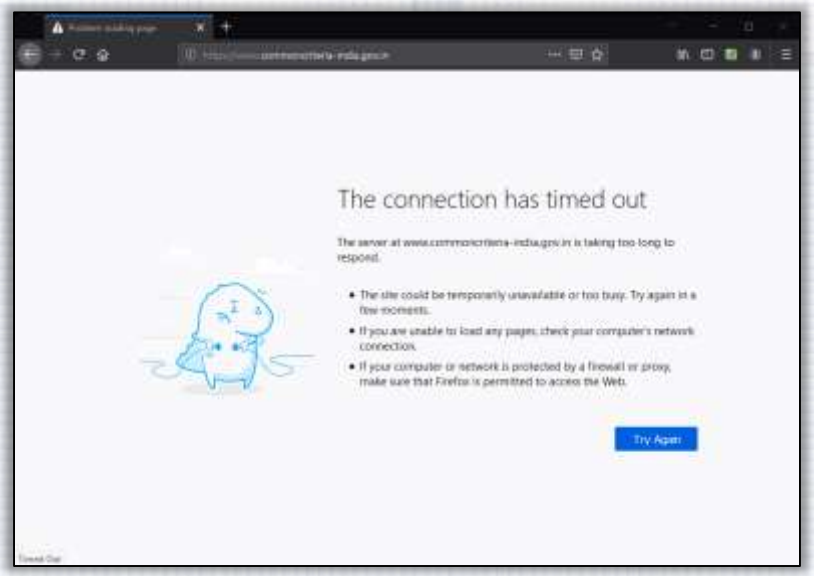

```
<div class="innerSearcher">
    <form method="get" id="searchForm" action="/search/totalsearch.do" >
        <input type="text" name="searchValue" class="button" placeholder="Search this site"/>
        <input type="hidden" id="csrf" name="csrf" value="YsyTsUHjFjCkIyE68BgdKm03xmRxcWGa" />
    </form>
</div>
<div class="innerMenu">
    <ul class="navWraper">
```

Difficulties while parsing

- Norwegian Certification Body does not have a list, instead they offer the possibility to directly search products.

Difficulties while parsing

- Indian CB site was down during weeks 38 and 39 of 2019 while we were trying get the latest data.
- Unfortunately, some data from the Indian CB could not be incorporated to our results.



Obtaining archived products from CBs

- French, Italian, Norwegian and Nederland CB sites offer the list of all certified products without distinction between archived and not archived.
- Australian and Singapore CBs have a small list of certified products, not categorizing archived items.
- German, Swedish, Korean and Japanese CBs had a separate list

Obtaining archived products from CBs

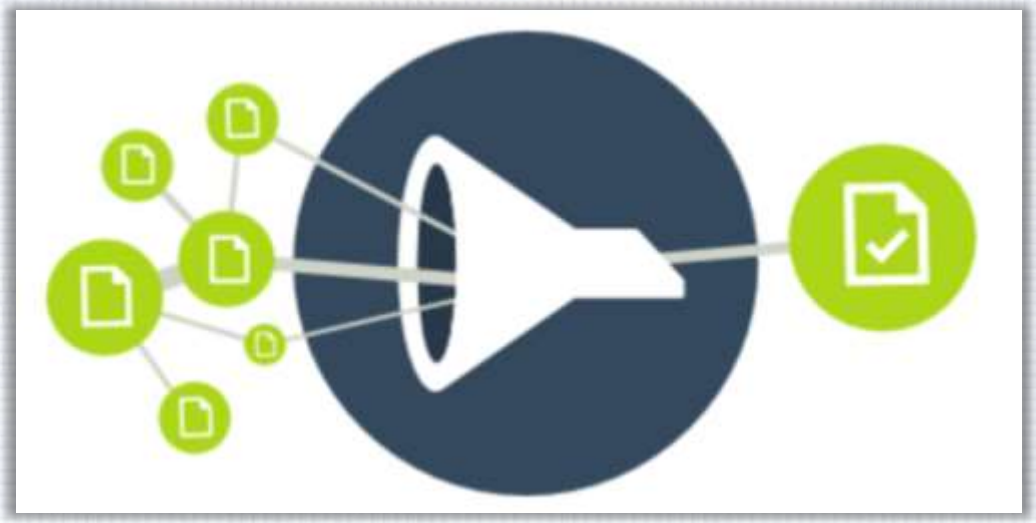
- Spanish CB doesn't include the list of archived products.
- NIAP has a table with the list of archived products, but there is no information besides the conformance claim and certification date.

Index

- Recap on CC Web Scrapper v1.0
- What's new in CC Web Scrapper v2.0
- Implementation and difficulties while parsing
- **Obtaining and correlating archived data from CBs**
- Technical conclusions
- Statistics
- Final conclusions

Correlation of data from CCPortal and CBs

- CB sites and CC Portal often provide slightly data with differences for the same certified product.
- The scrapper now includes a correlation logic for determining which items are the same and analyzing differences.



Correlation of data from CCPortal and CBs

- The scraper must be able to determine when two products, from CB site and CC Portal, are the same.
- CC Scrapper considers three cases for this:
  - Same normalized product name (collision risk)
    - Add certification date -> still risky, but acceptable
  - Same ST SHA -> some bytes change from CB version to CCPortal version
  - Same product name + Same ST SHA1: too rigid?

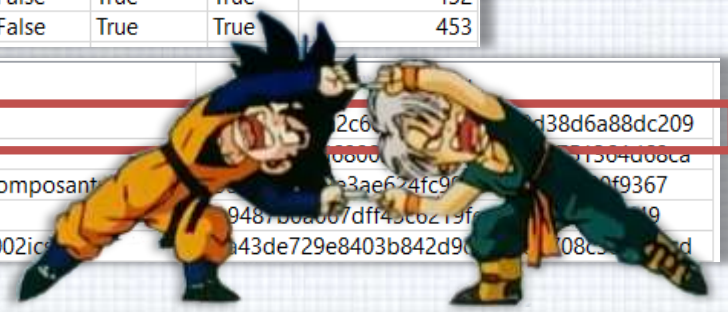
CB	Security Target SHA1	Duplicate	Scrapped from
BSI-DE	86ea7211c2cfd6d53628091753ea5194da5975ab	Hash	cc-cb
BSI-DE	3f0a3e20b6214f562ffd1cca04c13c5f1cb9c18b	Hash & name (Insurance Security Token Service (ISTS) Version 1.0)	cc-cb
MYCC-MY	15dbe30ffb3444f9d6c59cfae2bafdf03d6c1977		cb
BSI-DE	09192aef85a94e34a798c09a343d7355011fe575	Prodname (TCOS Residence Permit Card Version 1.1 Release 2-BAC/ SLE78CLX1440P)	cc-cb

Correlation of data from CCPortal and CBs

- For identifying unique products, each product found has its normalized name and hash calculated. Then, an unique product id is assigned.
- Each product parsed is associated with the unique product id of the related unique product.

	id	category	n_cat	p_index	archived	completed	fromcb	id_uniqprods
1541	2283	Multi-Function Devices	10	246	False	True	True	449
1542	1139	Other Devices and Systems – 203 Certified Products	12	26	False	True	False	450
1543	2098	Other Devices and Systems	13	54	False	False	True	450
1544	519	ICs, Smart Cards and Smart Card-Related Devices and S...	6	386	False	True	False	451
1545	2632	ICs, Smart Cards and Smart Card-Related Devices and S...	7	689	False	True	True	451
1546	1472	ICs, Smart Cards and Smart Card-Related Devices and S...	7	65	False	True	True	452
1547	1924	ICs, Smart Cards and Smart Card-Related Devices and S...	7	506	False	True	True	453

	id	product
450	450	sensor2185kitas40release10configurationoptions218520and218532
451	451	xsmartepassportv13r30ns3ct9kws3ct9kcs3ct9k9
452	452	applicationiasxlsurplateformejavacardenconfigurationouvertedelacartepucemultiappidv21surcomposant
453	453	microcontrleurscurist6nd7rvision4
454	454	taskalfa6002itaskalfa5002itaskalfa4002itaskalfa6002igtaskalfa5002igtaskalfa4002igkyoceracs6002ic



Correlation of data from CCPortal and CBs

- The same product when retrieved from CC Portal or CB websites may contain differences:
  - Slightly different product name / manufacturer
  - Abbreviated vs complete PP names
  - Conformance name: “PP Compliant” vs EAL-X
  - Inconsistent product categories with respect to CC Portal.
  - Different date formats.
- We had to create additional logic for these situations.

Correlation of data from CCPortal and CBs

- For some products, the scrapper did not found any matching item in two websites at all after a full run.
- This happens when a product is listed in the CB webpage but not in CC Portal.
- It also happens when an archived product listed in CC Portal from a CB that doesn't list duplicates in its site.

Index

- Recap on CC Web Scrapper v1.0
- What's new in CC Web Scrapper v2.0
- Implementation and difficulties while parsing
- Obtaining and correlating archived data from CBs
- **Technical conclusions**
- Statistics
- Final conclusions

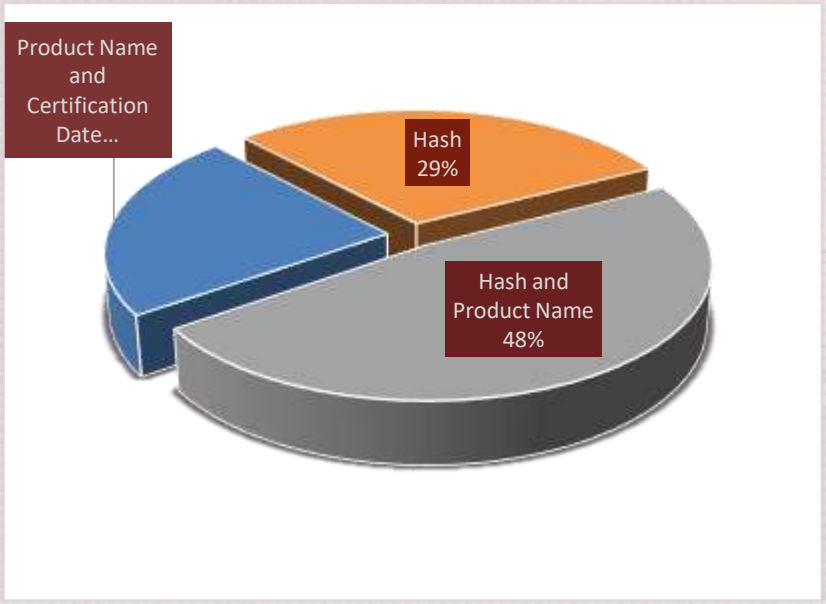
Technical conclusions

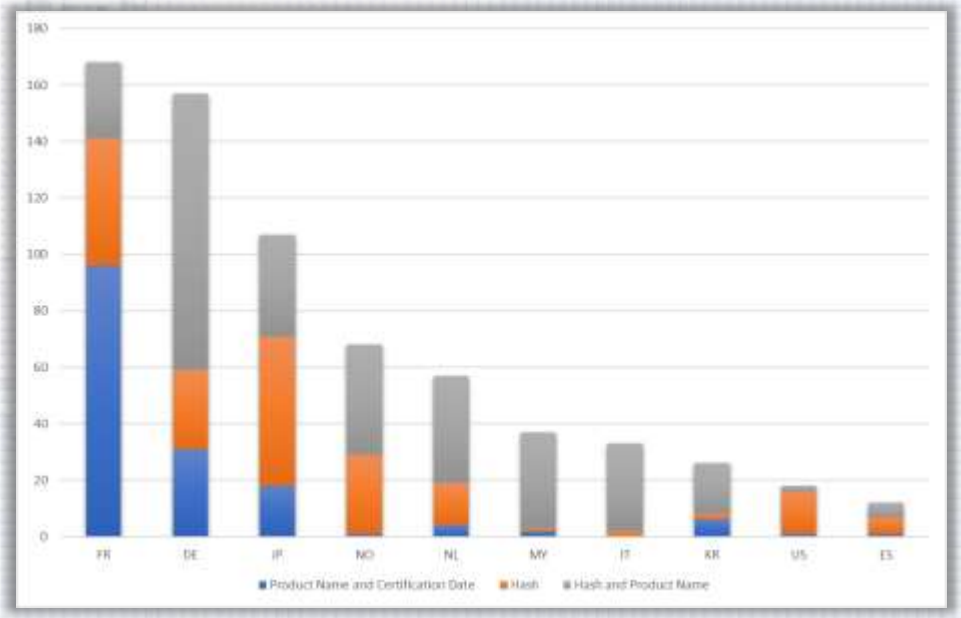
- Certified products are listed with a diverse and heterogeneous format in the different Certification Body websites.
  - Scrapping them is challenging.
  - Matching them is challenging.

- Some CBs don't list archived products; other CBs list them without tagging them as archived.

Technical conclusions

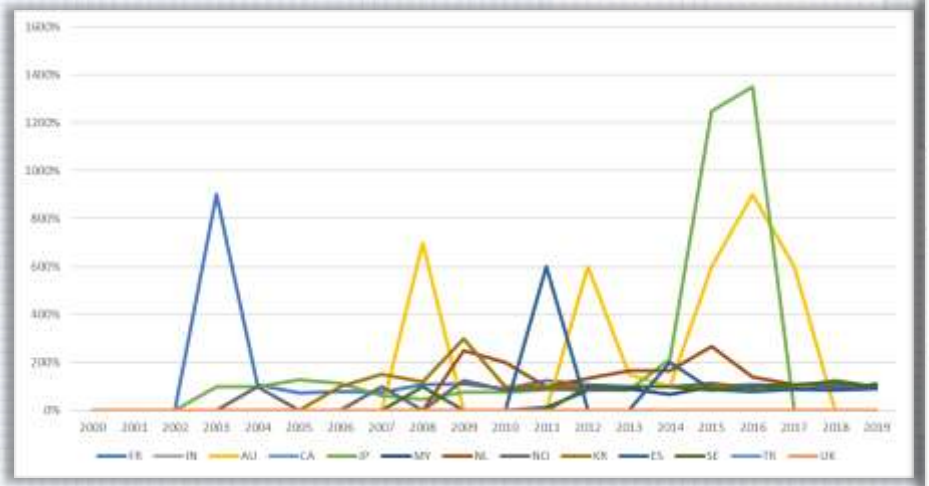
- The STs for the same product are not always the same identical file. Matching by product name + hash was required more than expected.

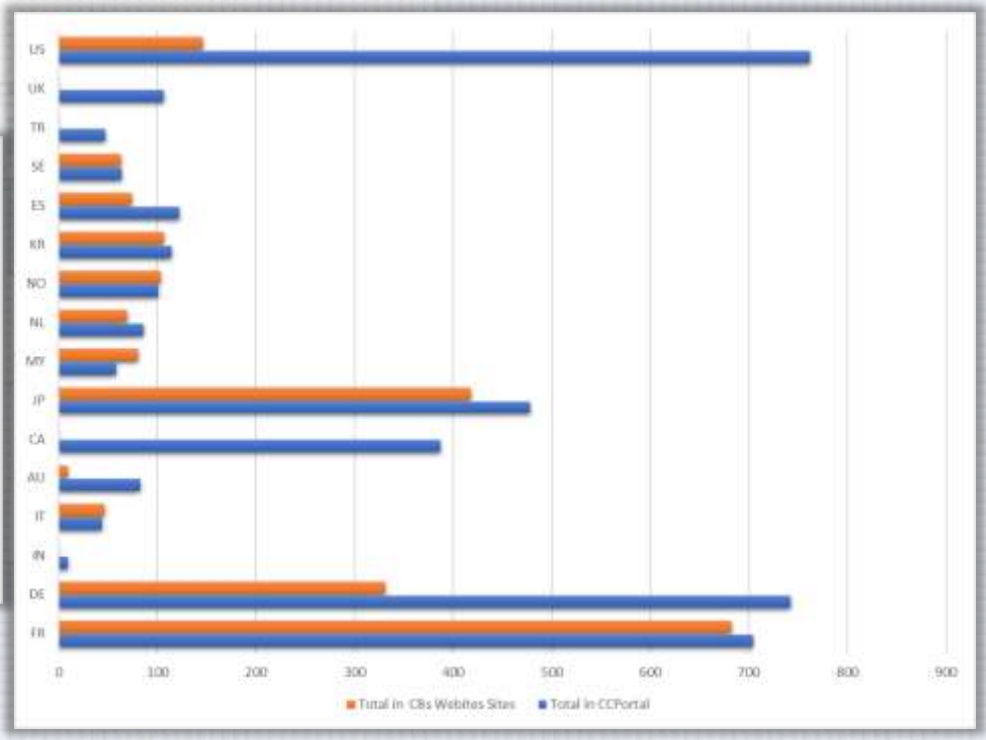




Technical conclusions

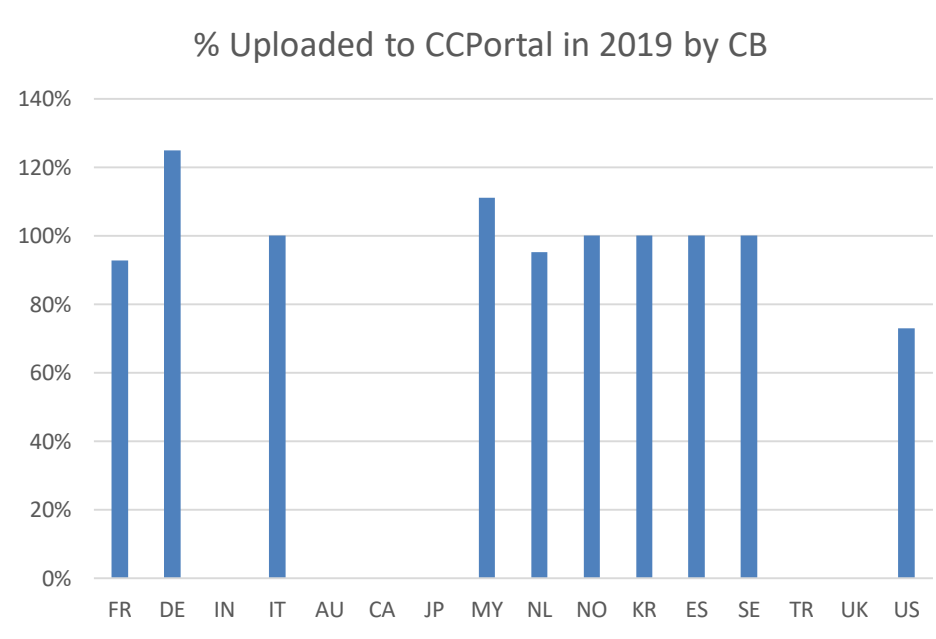
- (Crazy) Number of uploads from CB website to CC Portal





Technical conclusions

- We didn't achieve 1-to-1 correspondence because:



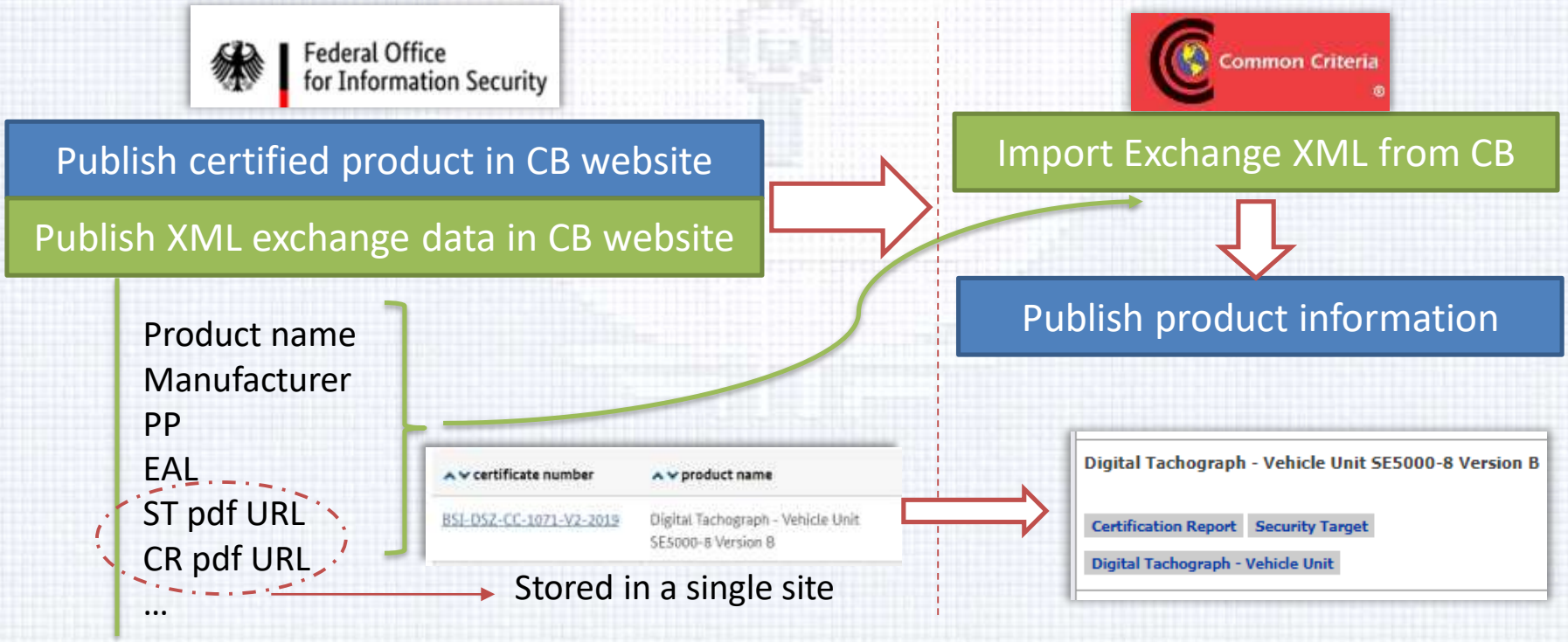
- a) Our algorithm didn't find the match (e.g. the ST is uploaded with different hash & different certification date)
- b) The product was still not uploaded when we execute the script
- c) The product was uploaded two times to CCPortal
- d) We miss something ☺
- AU/CA/IN/JP/TR/UK are only using CCPortal

Technical conclusions

- Is it possible to normalize this situation?
- Some CBs directly point to CC Portal. Is this the right approach?
- Even so, some information is entered manually in the website.

Technical conclusions

- We propose to design an **exchange protocol** (e.g. XML based) to publish the data once in CB / CC site and then synchronize from other sites.



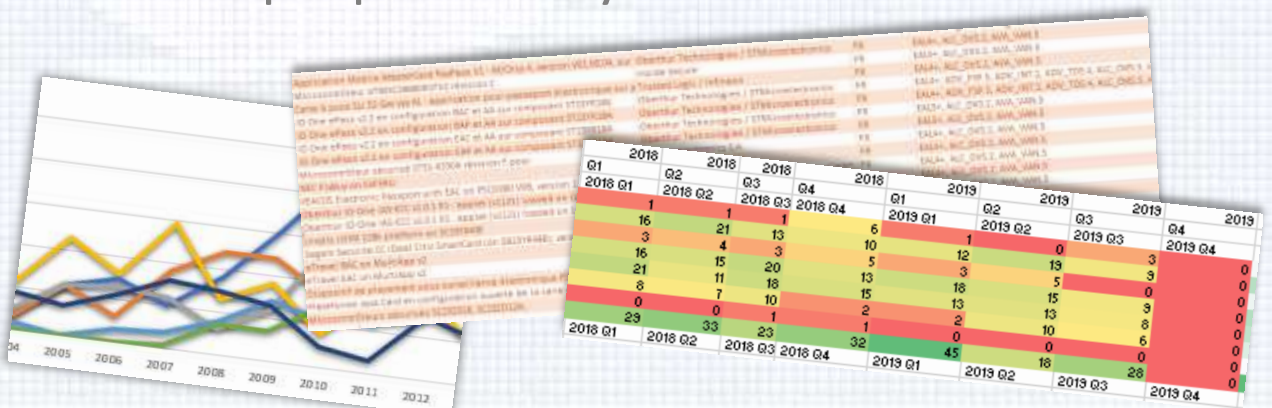
Index

- Recap on CC Web Scrapper v1.0
- What's new in CC Web Scrapper v2.0
- Implementation and difficulties while parsing
- Obtaining and correlating archived data from CBs
- Technical conclusions
- **Statistics**
- Final conclusions

Statistics – 2019 Q3

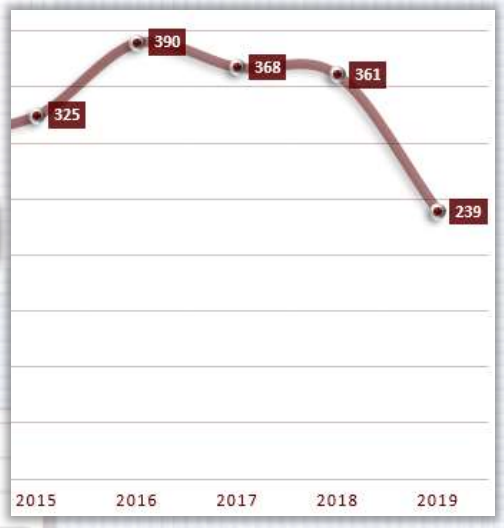
- Updated statistics for the period since the last ICCC will be presented now, to take a look at CC certification's health during recent times.

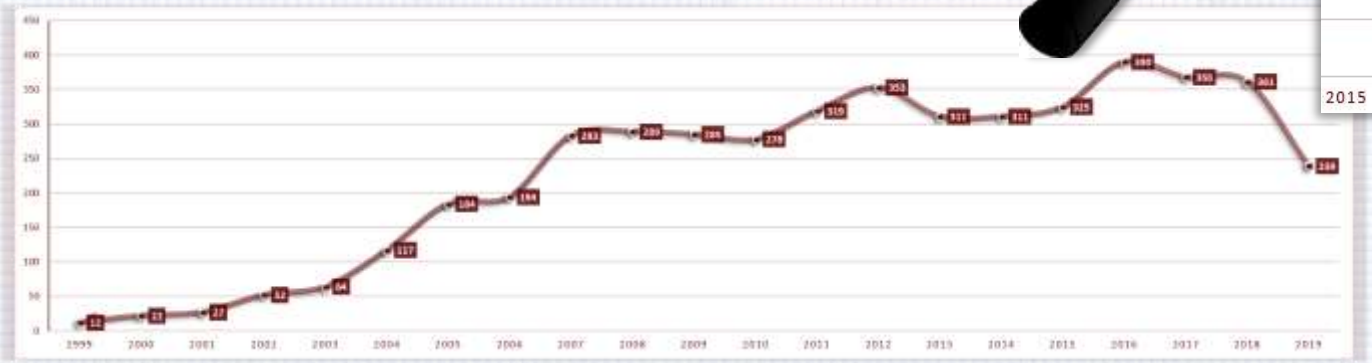
- The data is more complete and reliable now:
  - Information from the CBs websites was incorporated.
  - Archived products are considered for historical statistics purposes only.



Statistics – 2019 Q3

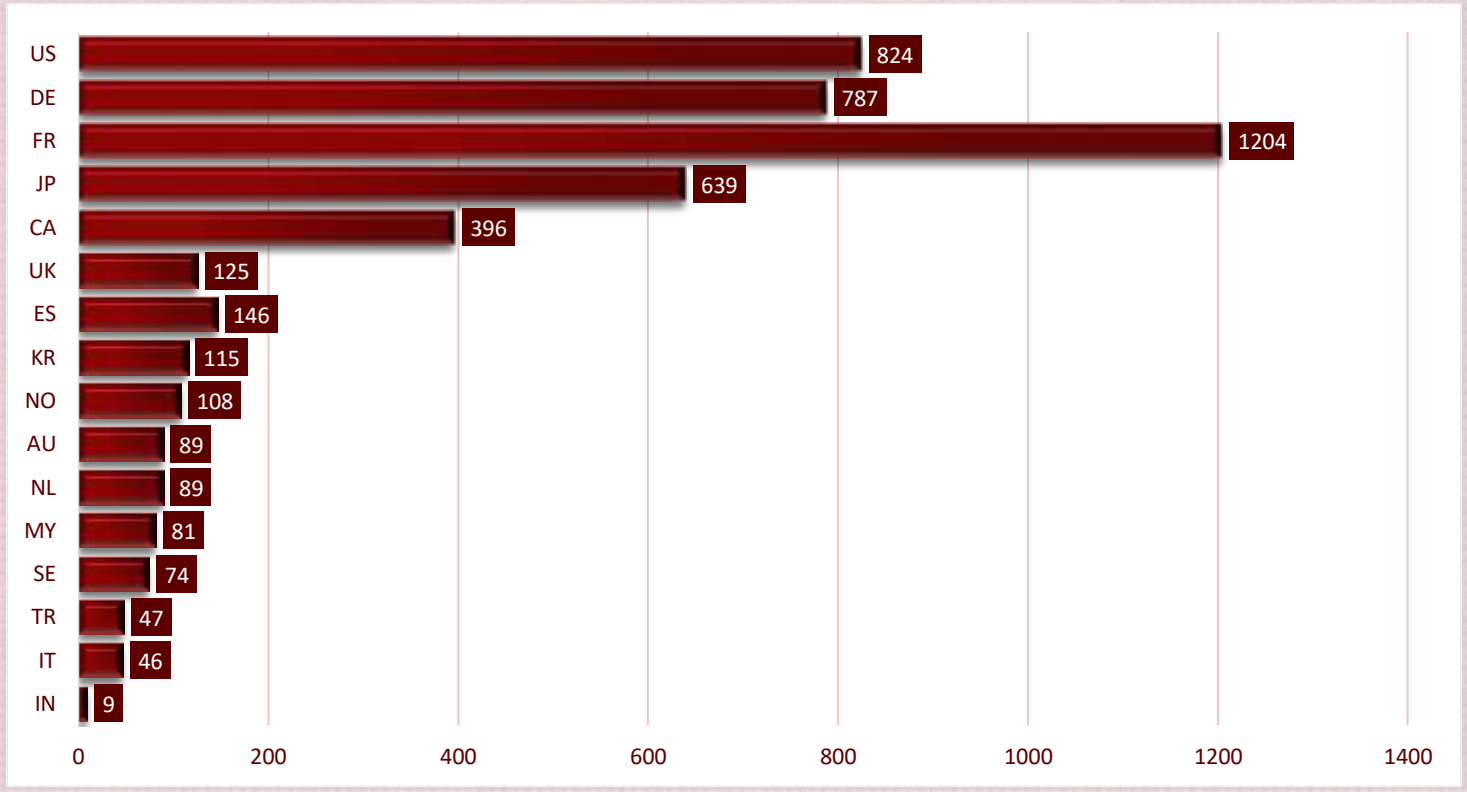
- Less certified products in 2019 (Q1-13) versus previous years, continuing a downward trend since 2016.
- 2019 may end near 2018 after Q4





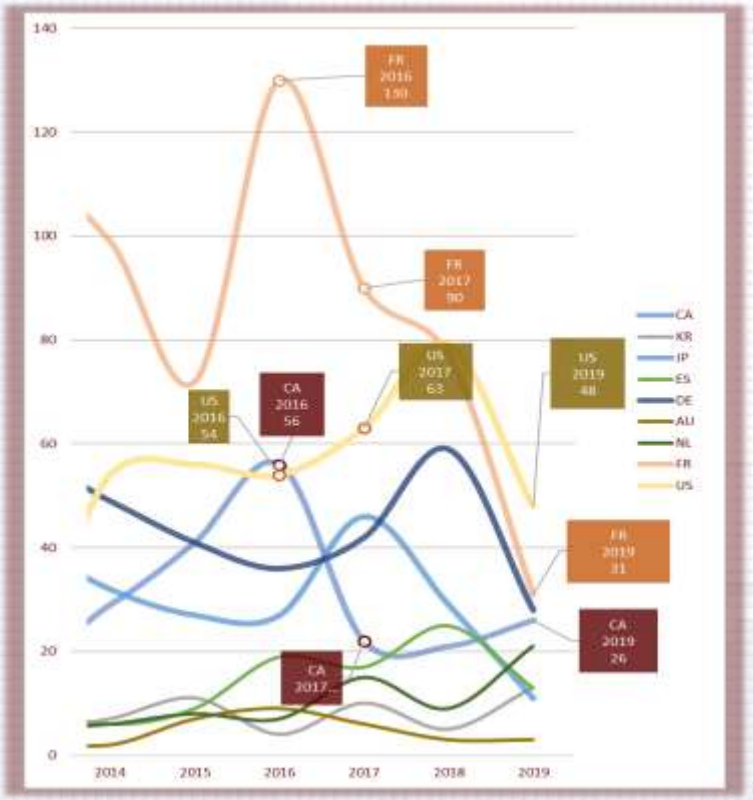
Statistics – 2019 Q3

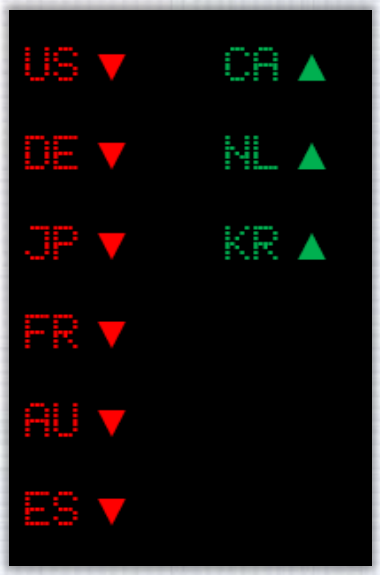
- Number of certifications per country, historical (archived included)



Statistics – 2019 Q3

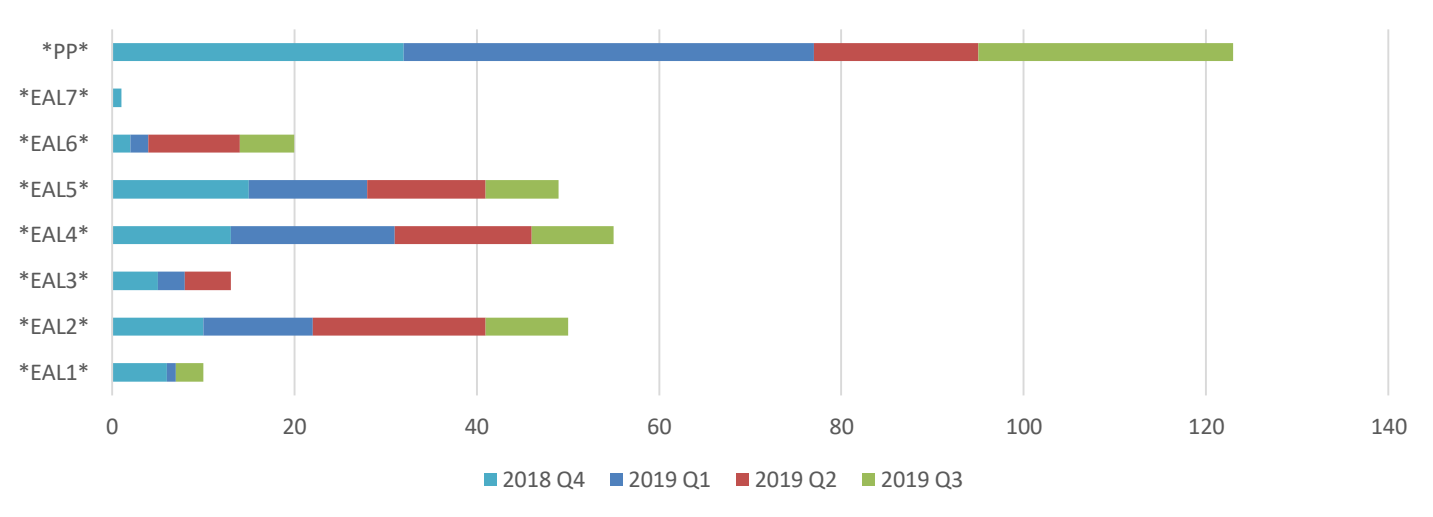
- Certifications per country, 5 year trend

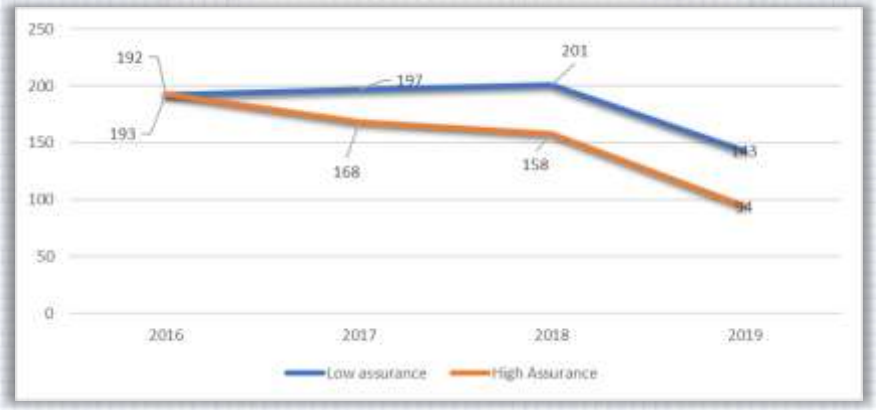




Statistics – 2019 Q3

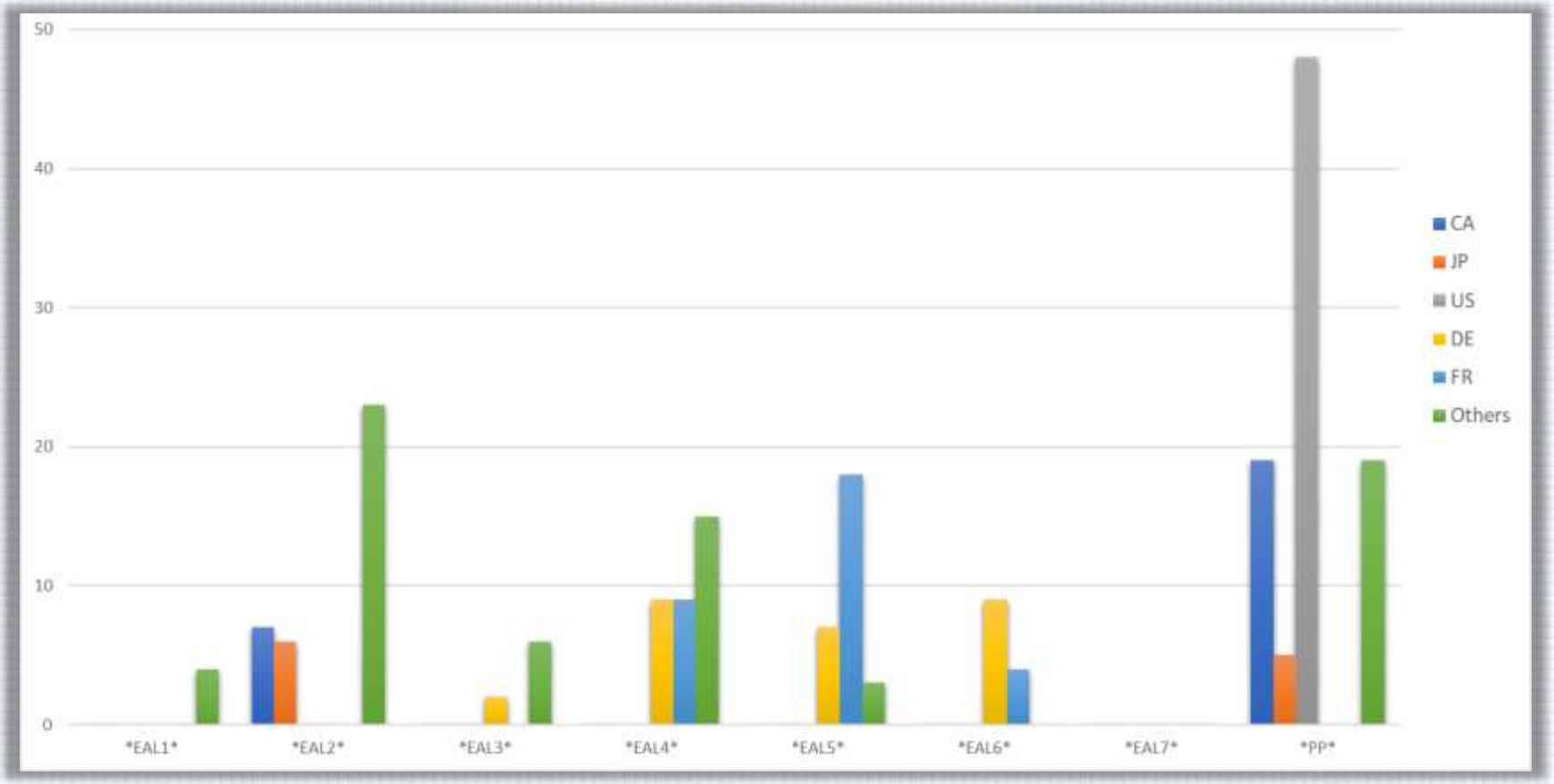
- Certified products compliance since last ICCC





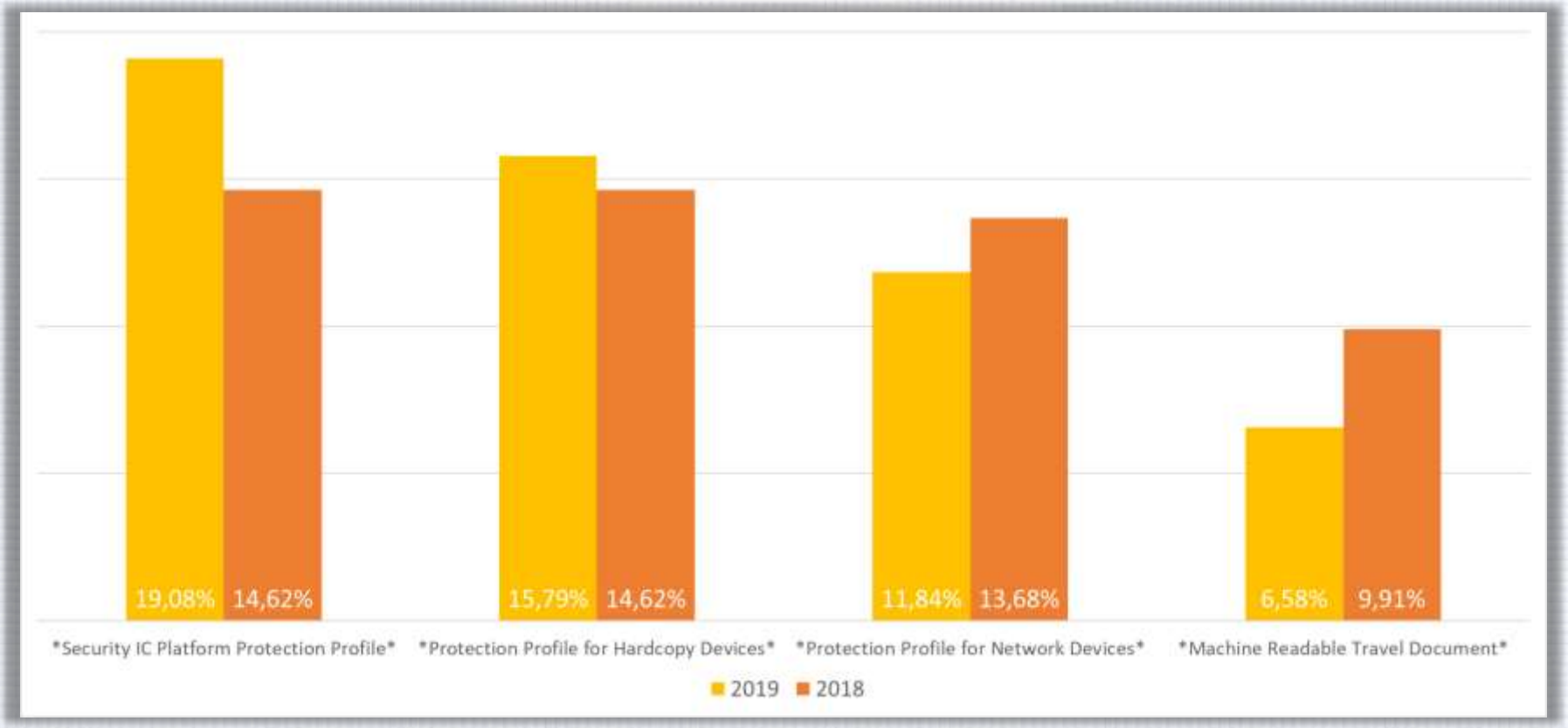
Statistics – 2019

- Product assurance level per country during 2019



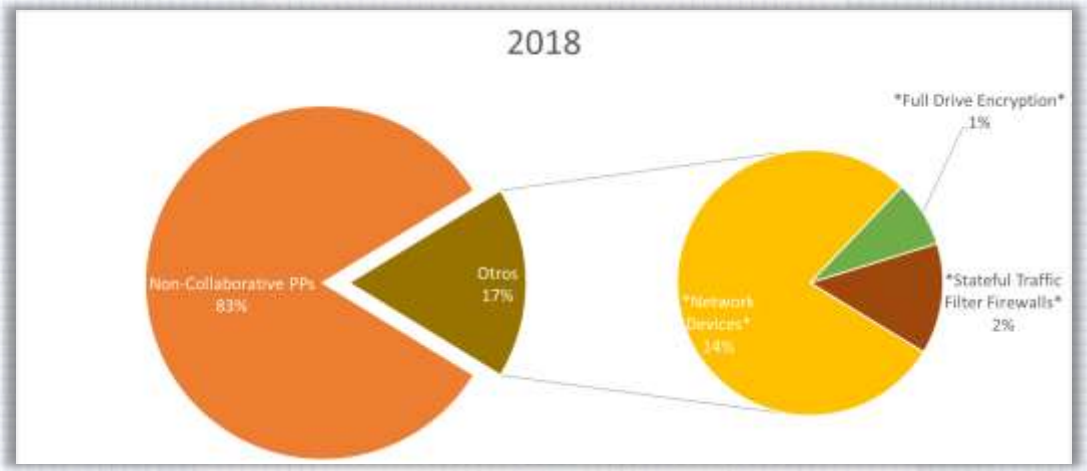
Statistics – 2019

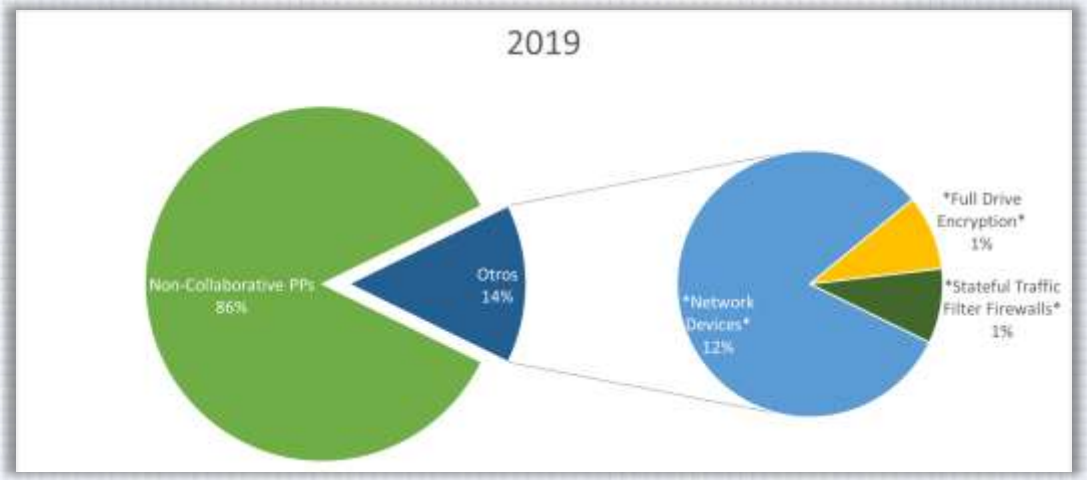
- Top PPs 2018-2019



Statistics – 2019 Q3

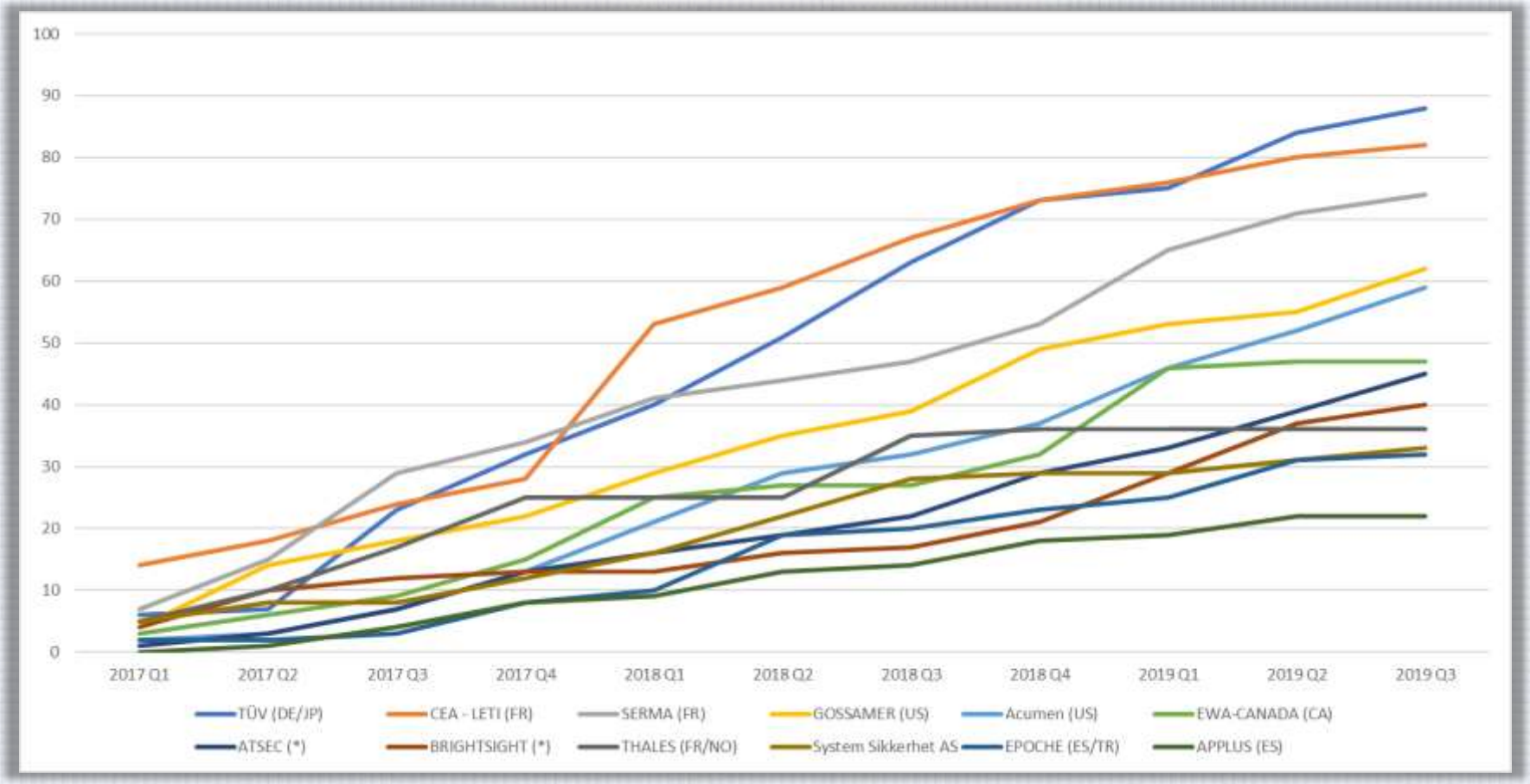
- Collaborative vs Non-Collaborative PPs 2018-2019





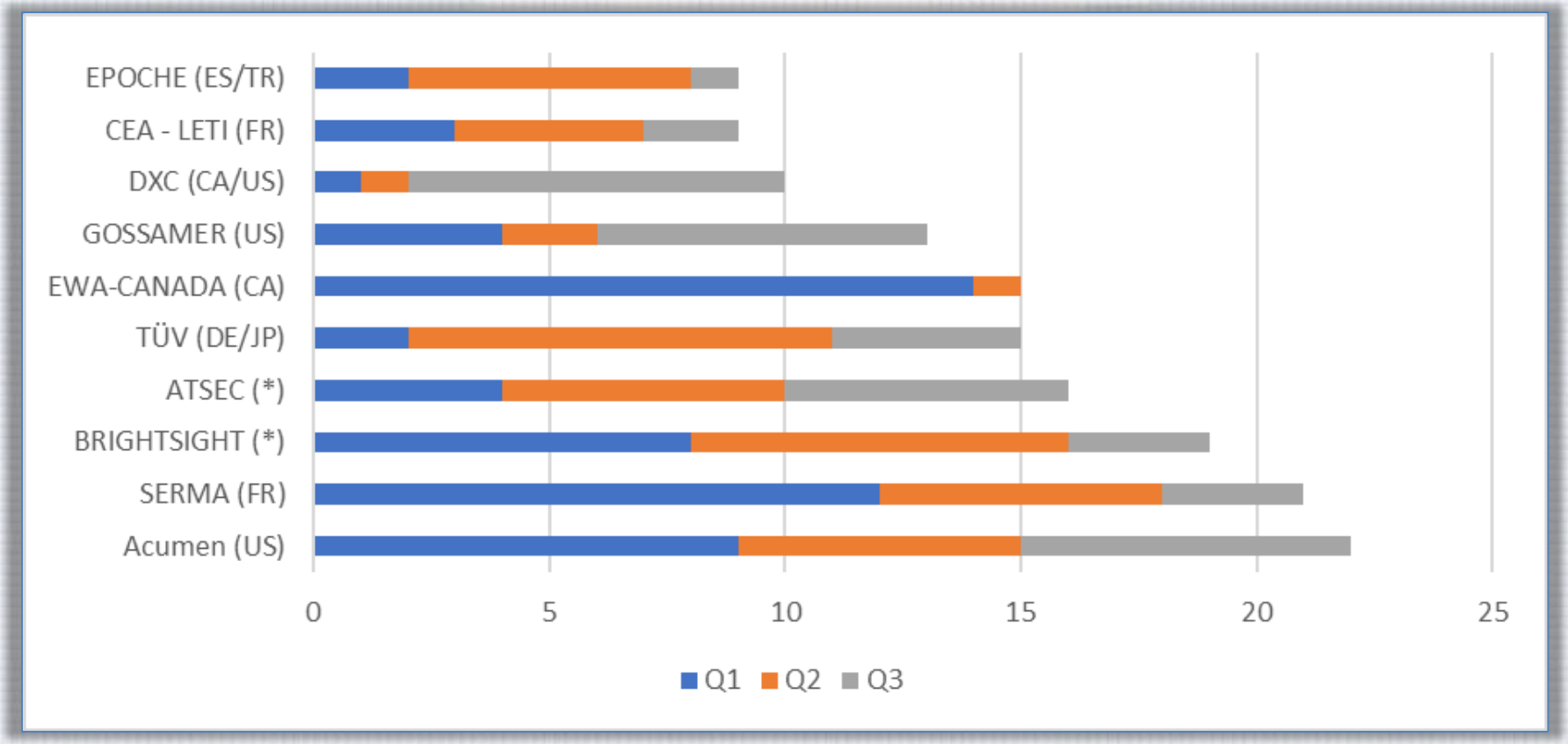
Statistics – 2019 Q3

- Certified products per Lab (3 years)



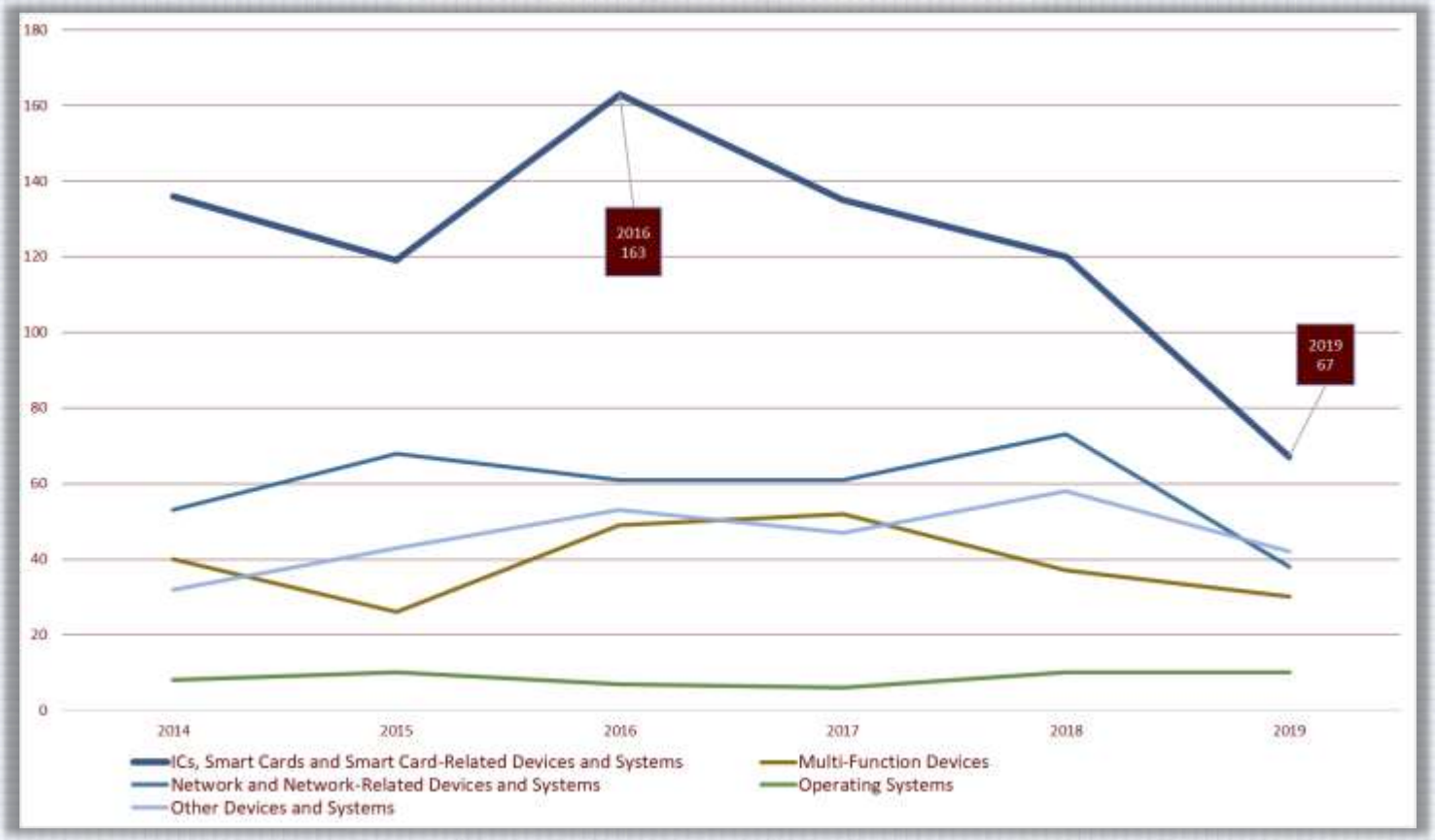
Statistics – 2019 Q3

- Top laboratories (2019 Q1-Q3)



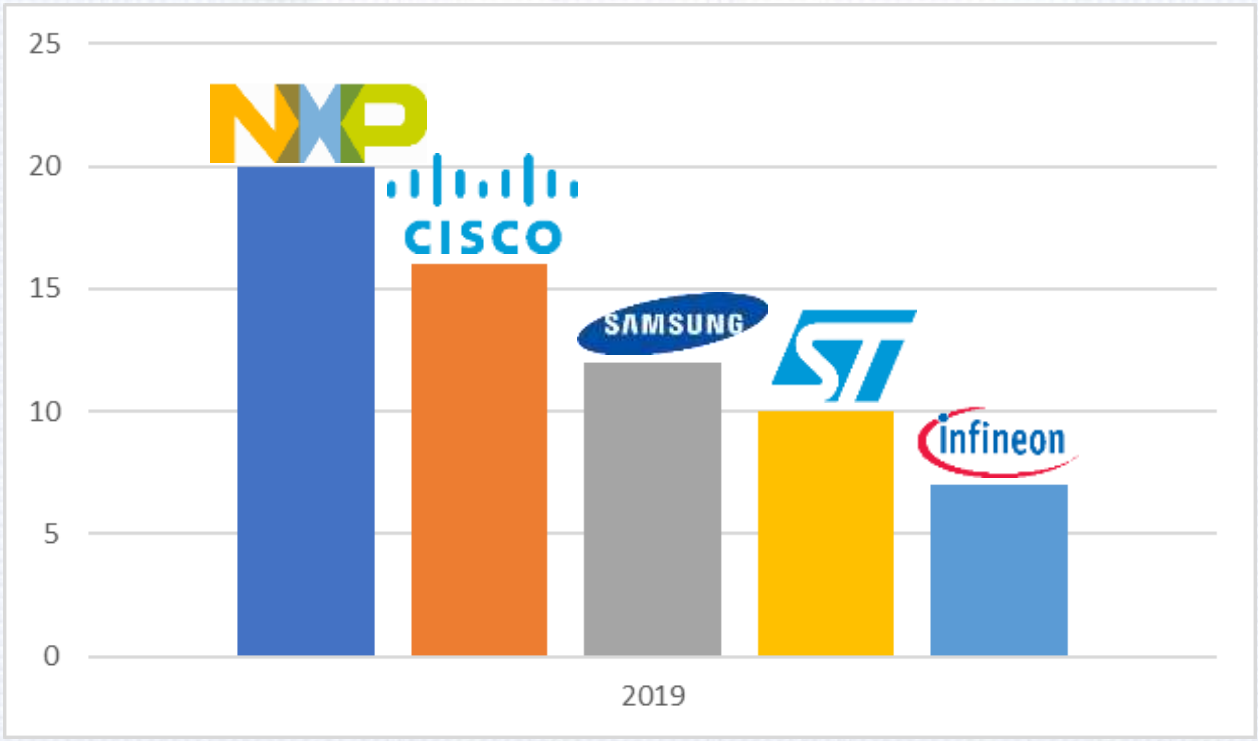
Statistics – 2019 Q3

- Evolution of top certified categories (5 years)



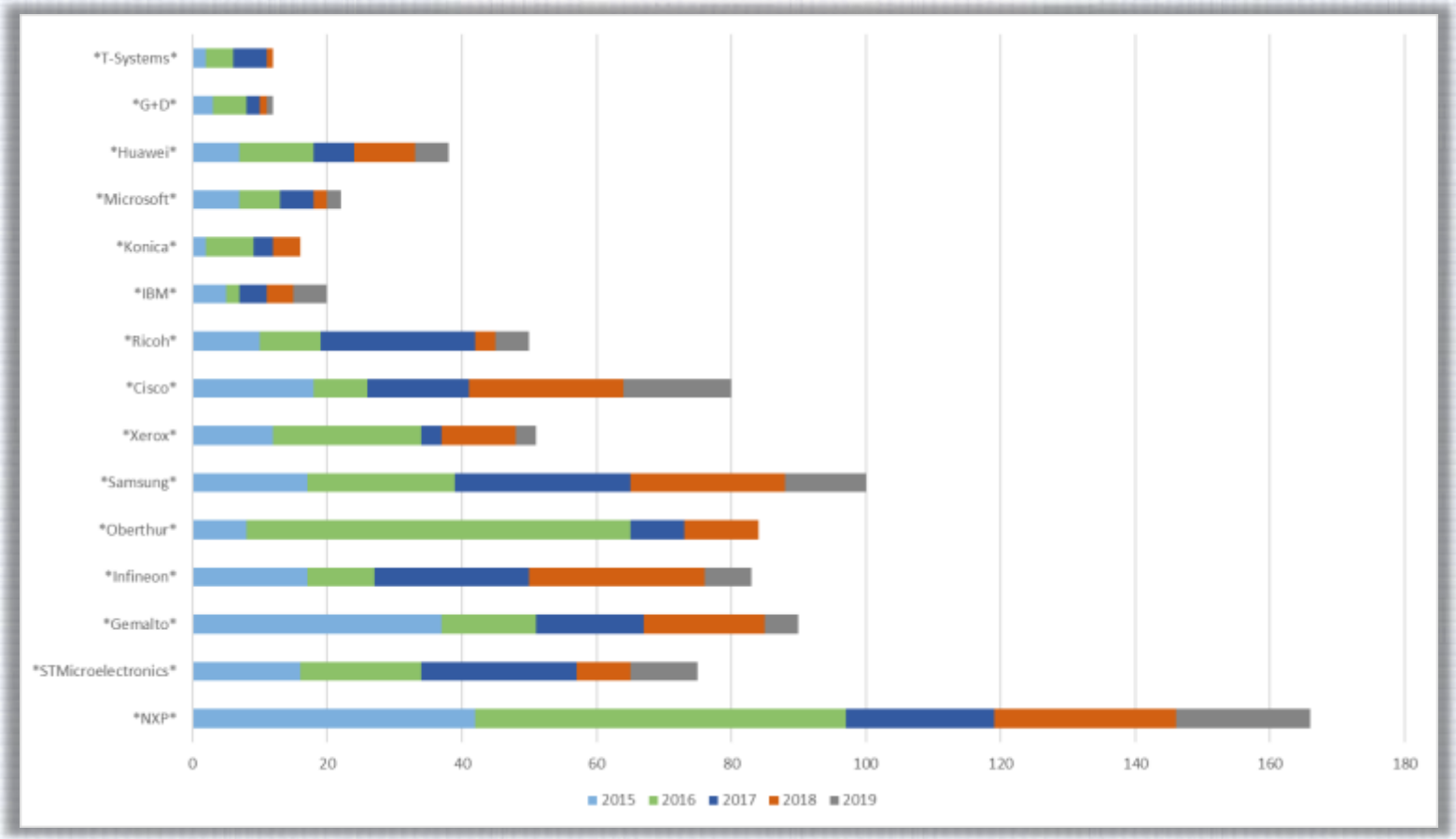
Statistics – 2019 Q3

- Top 5 manufactures of certified products (2019)



Statistics – 5 Years

- Manufacturer certifications per year (5Y)



Index

- Recap on CC Web Scrapper v1.0
- What's new in CC Web Scrapper v2.0
- Implementation and difficulties while parsing
- Obtaining and correlating archived data from CBs
- Technical conclusions
- Statistics
- **Final conclusions**

Conclusions

- There has been a decrease in the number of evaluations in the last 3 years compared to previous periods.
- The numbers of top certifying countries (DE, FR, US, JP) are in a downward trend.
- Smaller CBs are slightly rising their certification numbers.

Conclusions

- Since last ICCC, products certify mostly with PP compliance. High assurance certifications have been most frequent than low ones.
- NIAP has certified products almost exclusively with PP compliance during the last year.
- Other schemes certify roughly as much PP-compliant products as non PP-compliant ones.
- EAL2 has been by far the most frequent low assurance level during this period.

Conclusions

- During this year, most common PPs have been Security IC Platform PP, Hardcopy Devices PP, Network Devices cPP, and MRTD PP. This is a continuous trend in time.
- cPPs have been used by ~15% of PP-compliant certifications since 2018.
- Network Devices PP has become the most popular cPP, with a great difference over the rest.

Conclusions

- The traditional big labs (CEA-LETI, TÜV, SERMA, Acumen, Brightsight, Gossamer…) have been also on the top during the last years.
- epoche and DXC have entered the top 10 since last ICCC.
- Security ICs, Network Devices and Multi-function devices have been the most certified categories.
- NXP, CISCO, Samsung, ST Microelectronics and Infineon are 2019's top vendors, continuing their historical trend.

Datos de contacto

jtsec: Beyond IT Security

Avda. de la Constitución 20, Office 208

CP 18002 Granada – Spain

hello@jtsec.es

@jtsecES

www.jtsec.es

"Any fool can make something complicated. It takes a genius to make it simple." - Woody Guthrie