https://doi.org/10.2352/ISSN.2470-1173.2021.3.MOBMU-044


# Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 2: Censys

***Klaus Schwarz***[2,3]***, Reiner Creutzburg***[1,2]

[1] ***Technische Hochschule Brandenburg, Department of Informatics and Media, IT- and Media Forensics Lab, Magdeburger Str. 50, D-14770 Brandenburg, Germany***
***Email: creutzburg@th-brandenburg.de***

[2] ***SRH Berlin University of Applied Sciences, Berlin School of Technology, Ernst-Reuter-Platz 10, D-10587 Berlin, Germany***
***Email: klaus.schwarz@srh.de, reiner.creutzburg@srh.de***

[3] ***The University of Texas at San Antonio, College of Engineering, 1 UTSA Circle, San Antonio, TX 78249-0669, USA***

## Abstract

*Open-source technologies (OSINT) are becoming increasingly popular with investigative and government agencies, intelligence services, media companies, and corporations.*
*These OSINT technologies use sophisticated techniques and special tools to analyze the continually growing sources of information efficiently.*
*There is a great need for professional training and further education in this field worldwide.*
*After having already presented the overall structure of a professional training concept in this field in a previous paper [25], this series of articles offers individual further training modules for the worldwide standard state-of-the-art OSINT tools.*
*The modules presented here are suitable for a professional training program and an OSINT course in a bachelor's or master's computer science or cybersecurity study at a university.*
*In part 1 of a series of 4 articles, the OSINT tool RiskIQ PassivTotal [26] is introduced, and its application possibilities are explained using concrete examples. In this part 2 the OSINT tool Censys is explained [27]. Part 3 deals with Maltego [28] and Part 4 compares the 3 different tools of Part 1-3 [29].*

## Introduction and Motivation

CENSYS is often referred to as the most dangerous search engine in the world alongside SHODAN. CENSYS was developed by a group of researchers at the University of Michigan to make the Internet safer. Regular Internet-wide port scans across the entire public IPv4 address space can be used to identify vulnerable devices and networks and generate statistics on usage patterns of specific protocols or certificates. The results can be retrieved with an advanced full-text search or via an API [14].

## Censys – Accompanying theory

While similar search services of this kind focus on host discovery, CENSYS takes the path of performing complete protocol handshakes and analyzing the recorded data. This achieves a significantly higher hit rate without sacrificing accuracy.

The backend of CENSYS consists of the highly parallel application scanner ZGrab (part of the open-source project ZMap), which currently detects and analyzes numerous other application handshakes in addition to StartTLS, Heartbleed, and SSLv3 and makes them available as JSON objects. ZMap identifies the interesting hosts, and ZGrab initiates the handshakes and provides the corresponding structured data [13].
Using simple searches like

**443.https.heartbleed.vulnerable: true**

can be found, for example, by heartbleed vulnerable hosts. Filters can also be used to search for Poodle vulnerable hosts or outdated SHA-1 SSL Certificates [15].

With this new approach and the project's open-source nature, there is a transparent platform to uncover and detect global and local grievances [16]. All scanned data is available at **https://scans.io/**. The source code of ZMap and ZGrab is available on GitHub at **https://github.com/zmap**.

### First steps

CENSYS provides a complex web interface and detailed filtering options. The central part of the site consists of a search field (fig. 1) that divides the search into three separate areas:

- IPv4 host,
- Websites,
- Certificate.

The search is based on classic Boolean search queries. The following section explains the keyword system and the differences between the three search areas [10].

After entering the search line results, the results that were found based on this search appear. At the top right, one will find various options for viewing the results.

The first way to view it is the **Results** tab (Fig. 2), which outputs the search results as a list. (Here, for example, the category **IPv4 Hosts** was searched for **protocols:"80/http"**) (Fig. 3).

Next to it one will find the tab page **Map** (Fig. 4). By clicking on it, the output results are displayed on a map, as shown here:

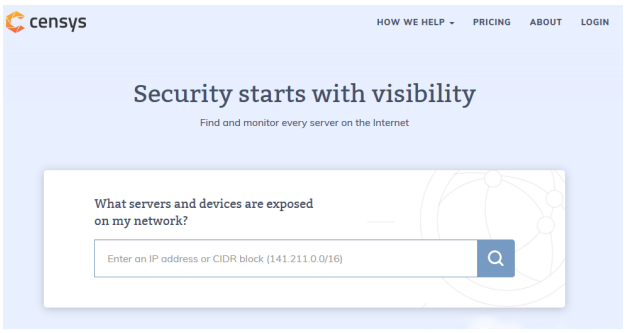



***Figure 1.*** CENSYS *search*

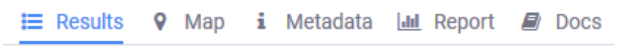



***Figure 2.*** ***Results*** *tab*

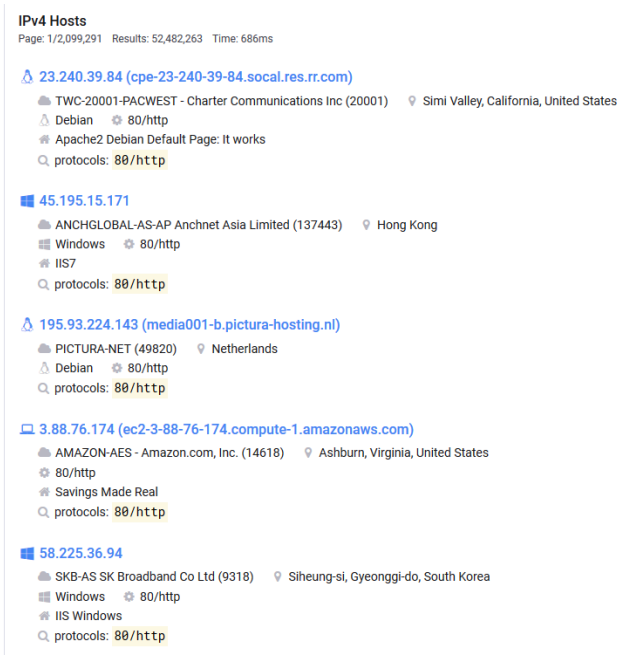



***Figure 3.*** *Search result for* **80/http**

Switching to the **Metadata** tab page (Fig. 5), one gets an overview of the result set.

The tab to the right of it has the name **Report**. This tool can generate a report on the breakdown of value present on the hosts returned by request.

The last and far-right tabs **Docs** contain important information about the syntax of CENSYS, examples of queries and data definitions, such as:

- timestamp (updated_at),
- tags (tags),


***Figure 4.*** *Output of the search on the map*

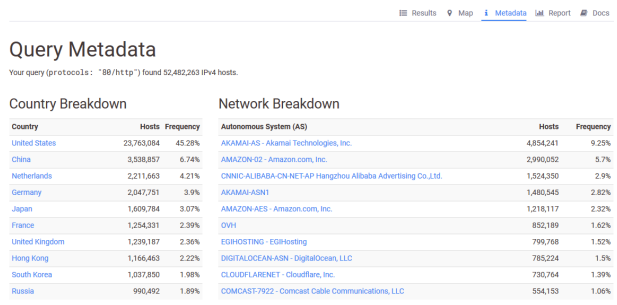



***Figure 5.*** *Result set*

- location data (location) and *autonomous system data (autonomous_system).*

In the left column there is a hint to **Quick Filters** and below that the number of **Protocols** and the so-called **Tags** (Fig. 6).

**Tags**

Tags are specific values that can be appended to some hosts. CENSYS includes heuristics that attempt to identify interesting hosts based on their banner responses and other factors [12]. Tags are then used to bundle hosts based on a specific factor. Tags include, for example:

- ‘‘camera’’ - Public cameras,
- "nas" - Network Attached Storage,
- "raspberry pi".

To view the complete list of tags, the URL **https://censys.io/ipv4/report** must be opened in the browser (Fig. 7). There the command **tags.raw** must be entered into the search field. The complete list should then be opened (Fig. 8).

**IPv4 hosts**

The primary purpose of CENSYS is to scan the IPv4 space, find open services, and collect the banners.

The Fig. 9 shows an example of an IPv4 results page. The host has two open services: **80/HTTP** and **443/HTTPS.** CENSYS


***Figure 6.*** *Quick filters*

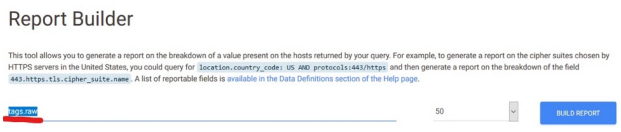



***Figure 7.*** *Input tags.raw*

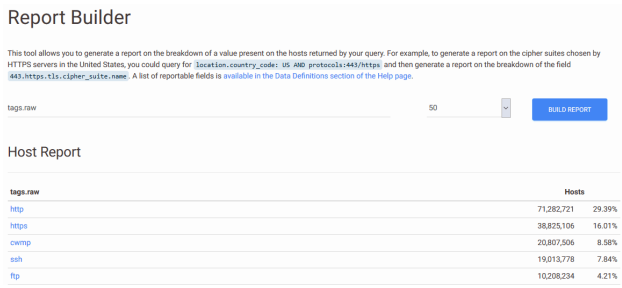



***Figure 8.*** *Complete list of tags*

allows filtering by protocol specific fields. For example, you can filter for the HTTP response status code for **80/http** and for a public key for **22/ssh** at the same time.

For example, to display a tag and a specific protocol:

**((195.37.1.142) AND tags.raw:"ssh") AND protocols:"22/ssh"**

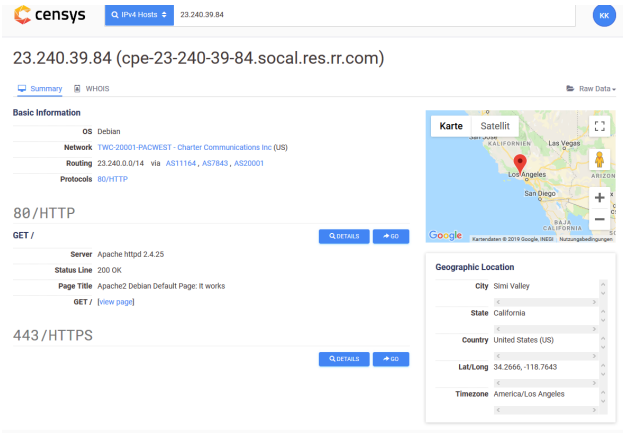



***Figure 9.*** *IPv4 results page*

**Further examples:**

Search queries are created as follows:

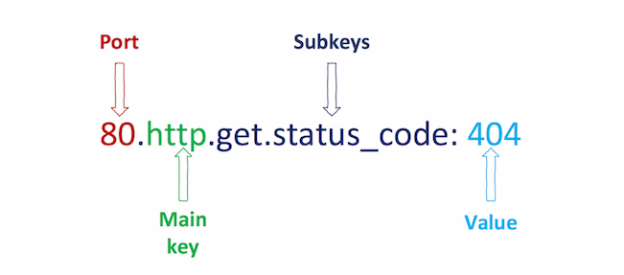



***Figure 10.*** *Structure of a search query*

The above query (Fig. 10) represents hosts that return status code 404 for **GET "/ "**. Several key-value pairs can be combined with **"AND"** or **"OR"**, e.g.:

Hosts who opened SSH and Telnet:
**ports:22 AND ports:23**

Hosts in Slovakia that are marked as camera:
**location.country_code: SK AND tags.raw: "camera"**

OpenSSH servers running on Debian all over Europe:
**22.ssh.v2.metadata.product: "OpenSSH"AND metadata.os: "Debian"AND**
**location.continent: "Europe"**

It is also possible to open the JSON output for a host that displays all key names. CENSYS also provides the ability to specify ranges.

Example: Hosts that have been updated since May and returned 5xx HTTP code.

**updated_at:[2018-05-01 TO *] AND 80.http.get.status_code:[500 TO 600]**

Regular expressions and placeholders are also possible. Example: **"name.first": "s.*y"**

### Websites

Websites offer mainly the same view as IPv4 hosts. CENSYS currently scans all domains from Alexa's top 1 million visited websites. Information about certificates and DNS (e.g., AXFR checks) is added to the usual host scan.

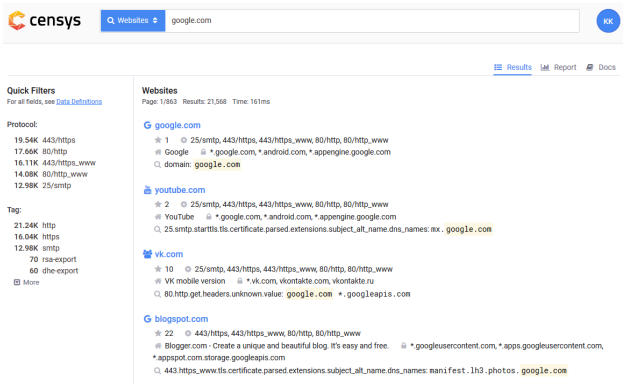



***Figure 11.*** *Search result for* ***google.com***

Search results for the website **google.com** were found here (Fig. 11).

One can choose between Results and Report as options for displaying the results.

### Certificates

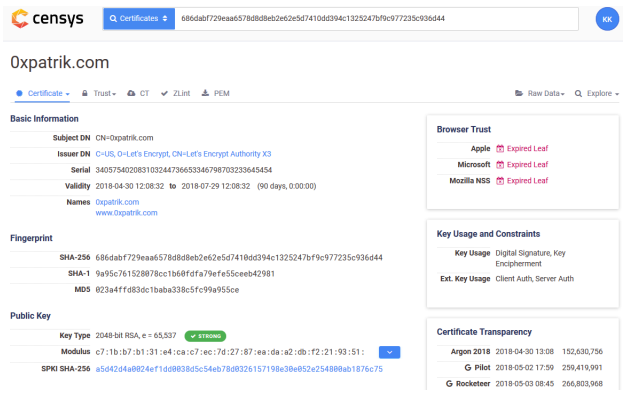



***Figure 12.*** *Result of a certificate search*

In CENSYS you can also search directly for certificates (Fig. 12). The only 20% of the certificates in CENSYS come from SSL scans in the IPv4 address space. The rest comes directly from the Certificate Transparency (CT) protocols.

The Certificate Transparency Project was created to get an overview of the CAs issue certificates. The process looks like this (Fig. 13):

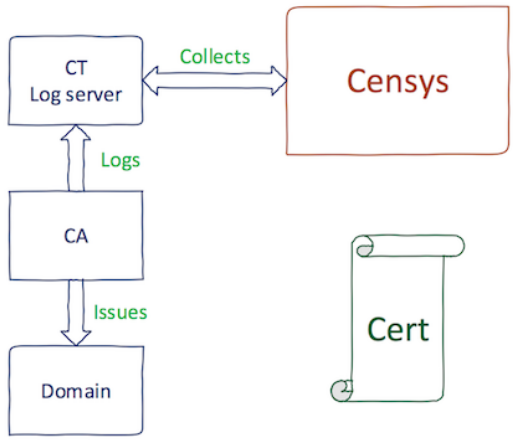



***Figure 13.*** *Certificate Transparency Project*

This means that in CENSYS the certificates are made available immediately after issue (almost in real-time) [11]. CENSYS does not have to rely on IPv4 scans to find certificates.

Especially when searching for certificates CENSYS offers one of the best services currently available.

CENSYS offers several new keys for the certificate search, such as

- Certificate Fields (parsed),
- Certificate Metadata (metadata),
- Trust Chain (validation),
- Certificate Transparency (ct.)

To view the complete list of certificates, the URL **https://censys.io/certificates** must be opened in the browser. A list of certificates should appear (Fig. 14).

**Examples:**

Certificates for www.example.com:
**parsed.names:"www.example.com"**

Certificates that are valid from Apple, but not from Mozilla NSS:
**validation.apple.valid:true AND validation.nss.valid:false**

Certificates issued by Symantec that were seen during the SSL search:
**parsed.issuer.organization.raw:"Symantec Corporation"AND**
**metadata.seen_in_scan:true**

Certificates issued by Let's Encrypt in 2018:

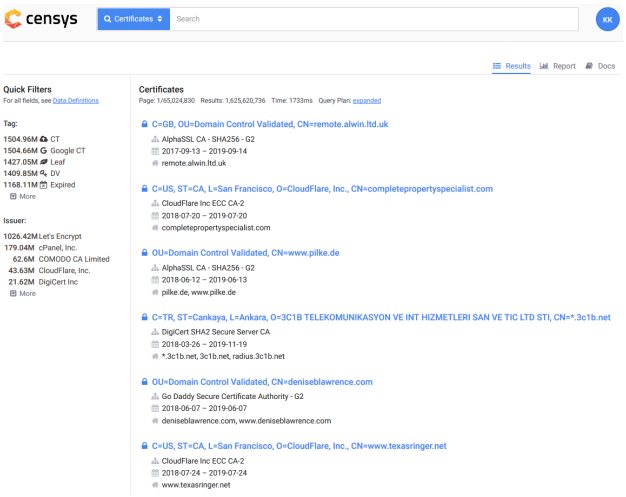



***Figure 14.*** *List of certificates*

**parsed.issuer.organization.raw:"Let's Encrypt"AND parsed.validity.start:[2018-01-01 TO *]**

CENSYS also offers interesting pivot/discovery techniques. Thus, several common factors can be searched for. To open it, click on the **Explore** tab in the upper right corner (Fig. 15).

## Censys API

### Exercise 1 (Obtaining the CENSYS API key)

CENSYS *also offers an API interface. This can be set up as follows.*

1. *Please visit the following page first:* **https://censys.io/account** *(Fig. 16).*
2. *Now select the item* ***API*** *(Fig. 17):*
3. *Make a note of your API ID and the corresponding secret. The API is a simple REST API. This means that calls can be made using fixed addresses. A simple client is even possible with the Linux console. A simple Python client looks like this (Fig. 18):*

### Exercise 2 (Setup of CENSYS API client)

*Of course,* CENSYS *also offers a ready-made client for your own use and further development. It is a Python client. The developer page of the Python client is available at the following URL:*

**https://github.com/censys/censys-python**

*However, it is constrained and can instead be regarded as a starting point for its own development. In this exercise, you will, therefore, use a third-party implementation of the client. This is available at the following address:*

**https://github.com/gelim/censys**

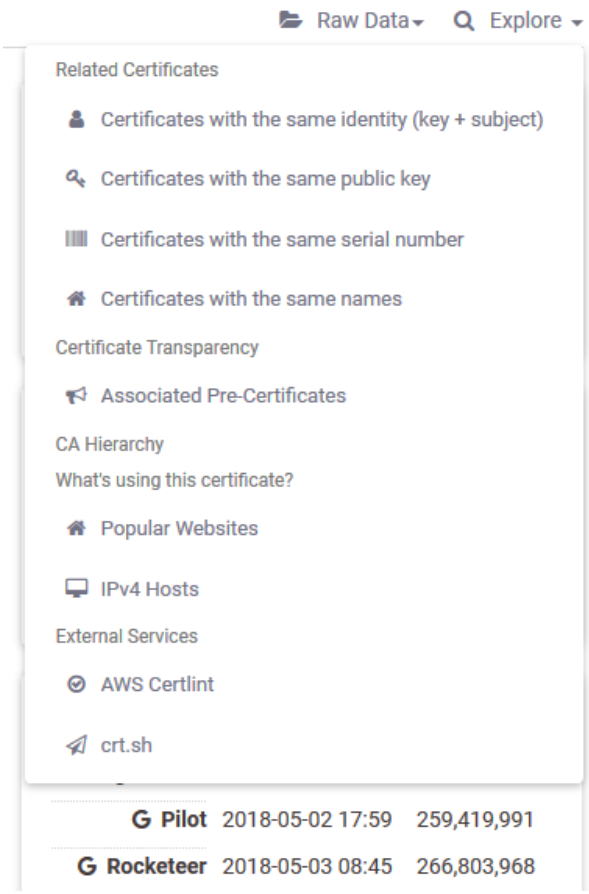



***Figure 15.*** *Options of Explore*

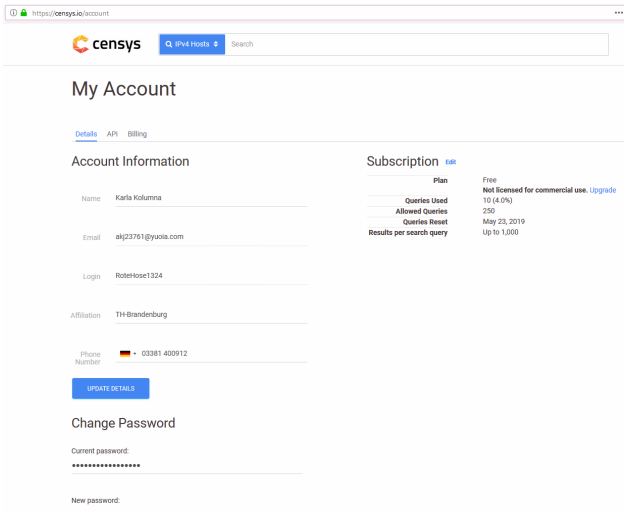



***Figure 16.*** CENSYS *account*

*The client offers numerous possibilities. The syntax of the*

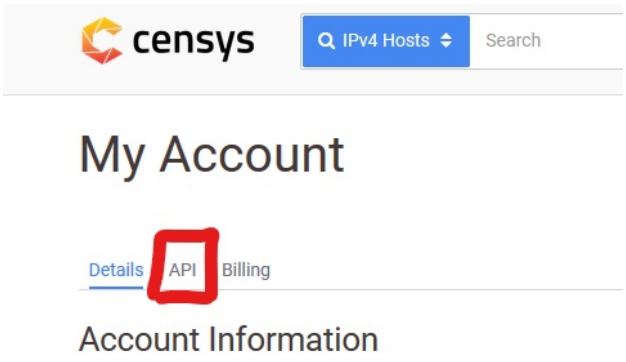



***Figure 17.*** CENSYS *API*


***Figure 18.*** *Simple Python client*

*client is as follows (Fig. 19):*

*The client can be set up relatively easily in the previously used virtual machine.*

1. *Open a terminal window and enter the following commands, where your API data must replace xxxx and yyyy:*

   **(a) git clone https://github.com/gelim/censys.git**
   **(b) cd censys**
   **(c) pip install -r requirements.txt**
   **(d) export CENSYS_API_ID=xxxx;**
   **export CENSYS_API_SECRET=yyyy**
   **(e) echo "export CENSYS_API_ID=xxxx"**
   **> > ~/. zshrc;**
   **echo "export CENSYS_API_SECRET=yyyy"**
   **> > ~/. zshrc;**

2. *After entering and successfully running the commands, you should now be in the program directory. Start the program with the following command:*


***Figure 19.*** *Syntax Python client*

**./censys_io.py -h**

3. *Now, you should be able to read the output as shown in the syntax image above. After the successful setup, the client can be used.*

## Selected console query examples

### Example 1

You want to count how many servers have SAP in their server header. The following request is made (Fig. 20):

**./censys_io.py -S SAP -count**

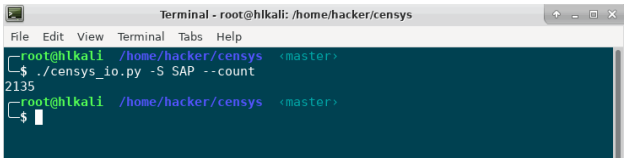



***Figure 20.*** *Counting with the* CENSYS *API client*

### Example 2

A search for hosts that have SSH and Telnet open is given by (Fig. 21):

**./censys_io.py ports:22 AND ports:23**

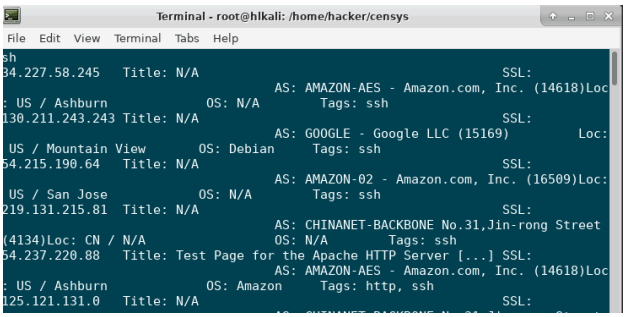



***Figure 21.*** *Output a search for special ports with the* CENSYS*-API Client*

**Example 3**

Searching for OpenSSH servers running on Debian across Europe (Fig. 22):

**./censys_io.py 22.ssh.v2.metadata.product:"OpenSSH" AND metadata.os:"Debian" AND location.continent:"Europe"**

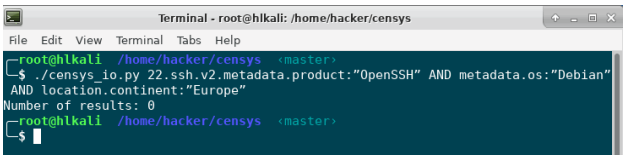



***Figure 22.*** *Output of a search for special servers with the* CENSYS*-API client*

**Rate limits**

The rate limits can be viewed in the account settings. Here is an example for rate limits of a CENSYS account (Fig. 23):

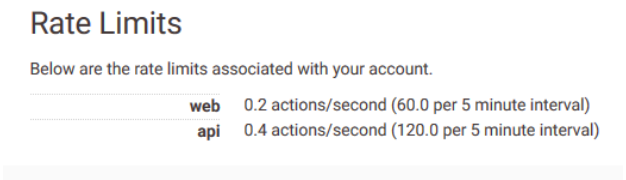



***Figure 23.*** *Rate limits*

## Summary

In today's dynamic IT environment, many organizations struggle to find every system and application at risk before the attackers do [9]. CENSYS empowers defenders with the automated visibility they need to truly understand and to get ahead of these risks, enabling even small security teams to have an outsized impact. For this reason, it is important to have trained operators who have the necessary expertise and can use this tool effectively and profitably.

The course created in this work gives participants a comprehensive overview of the topic of Open Source Intelligence. It allows them to work independently with the newly taught tools, collect data from freely available sources, and aggregate it into investigation results. For this purpose, the tasks were designed for several laboratory exercises.

## References


[1] Mercado, Stephen C. "Sailing the Sea of OSINT in the Information Age." Secret intelligence: A reader 78 (2009).

[2] Best Jr, Richard A., and Alfred Cumming. "Open source intelligence (OSINT): issues for congress." December 5 (2007): 28.

[3] Glassman, Michael, and Min Ju Kang. "Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT)." Computers in Human Behavior 28.2 (2012): 673-682.

[4] Quick, Darren, and Kim-Kwang Raymond Choo. "Digital forensic intelligence: Data subsets and Open Source Intelligence (DFINT+ OSINT): A timely and cohesive mix." Future Generation Computer Systems 78 (2018): 558-567.

[5] Williams, Heather J., and Ilana Blum. "Defining second generation open source intelligence (OSINT) for the defense enterprise." RAND Corporation Santa Monica United States, 2018.

[6] Benes, Libor. "OSINT, new technologies, education: Expanding opportunities and threats. A new paradigm." Journal of Strategic Security 6.3 (2013): 22-37.

[7] Schaurer, Florian, and Jan Störger. "The evolution of open source intelligence (OSINT)." Journal of US Intelligence Studies 19.3 (2013): 53-56.

[8] Pringle, Robert W. "The limits of OSINT: Diagnosing the Soviet media, 1985-1989." International Journal of Intelligence and CounterIntelligence 16.2 (2003): 280-289.

[9] Gibson, Helen. "Acquisition and preparation of data for OSINT investigations." Open Source Intelligence Investigation. Springer, Cham, 2016. 69-93.

[10] Carroll, Jami M. "OSINT Analysis using Adaptive Resonance Theory for Conterterrorism Warnings." Artificial Intelligence and Applications. 2005.

[11] Best, Clive. "OSINT, the Internet and Privacy." EISIC. 2012.

[12] Casanovas, Pompeu. "Cyber warfare and organised crime. A regulatory model and meta-model for open source intelligence (OSINT)." Ethics and Policies for Cyber Operations. Springer, Cham, 2017. 139-167.

[13] Layton, Robert, and Paul A. Watters. "Automating Open Source Intelligence: Algorithms for OSINT." Syngress, 2015.

[14] Steele, Robert David. "Open Source Intelligence (OSINT)."

[15] Berghel, Hal. "Robert David Steele on OSINT." Computer 47.7 (2014): 76-81.

[16] Weaver, Greg S. "Open Source Intelligence (OSINT)." The

Police and the Military: Future Challenges and Opportunities in Public Safety 4.

[17] Revell, Quentin, Tom Smith, and Robert Stacey. "Tools for OSINT-Based Investigations." Open Source Intelligence Investigation. Springer, Cham, 2016. 153-165.

[18] Kalpakis, George, et al. "OSINT and the Dark Web." Open Source Intelligence Investigation. Springer, Cham, 2016. 111-132.

[19] Tabatabaei, Fahimeh, and Douglas Wells. "OSINT in the Context of Cyber-Security." Open Source Intelligence Investigation. Springer, Cham, 2016. 213-231.

[20] Danda, Matthew. "Open Source Intelligence and Cybersecurity." (2019).

[21] Steele, Robert D. "1997 OSINT What Is It – Why Is It Important to the Military (White Paper)." Academia.edu **www.academia.edu/9817888/1997_OSINT_What_Is_It_Why_Is_It_Important_to_the_Military_White_Paper_**.

[22] "Social Media Prisma 2017/2018, Ethority, **ethority.de/social-media-prisma/**.

[23] Mohsin, Maryam, et al. "10 Social Media Statistics You Need to Know in 2020 [Infographic]." Oberlo, Oberlo, 15 Jan. 2020, **www.oberlo.com/blog/social-media-marketing-statistics**.

[24] Tenzer: "Daten - Volumen Der Weltweit Generierten Daten 2025." Statista, Statista, 13 Feb. 2020, **de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen/**.

[25] Schwarz, Klaus; Franziska Schwarz, Reiner Creutzburg: "Conception and implementation of professional laboratory exercises in the field of open source intelligence (OSINT)". Proceed. Electronic Imaging Symposium 2020 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2020, **https://doi.org/10.2352/ISSN.2470-1173.2020.3.MOBMU-278**

[26] Schwarz, Klaus; Reiner Creutzburg: "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 1: RiskIQ PassiveTotal". Proceed. Electronic Imaging Symposium 2021 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2021 (in print)

[27] Schwarz, Klaus; Reiner Creutzburg: "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 2: Censys". Proceed. Electronic Imaging Symposium 2021 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2021 (in print)

[28] Schwarz, Klaus; Reiner Creutzburg: "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 3: Maltego". Proceed. Electronic Imaging Symposium 2021 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2021 (in print)

[29] Schwarz, Klaus; Reiner Creutzburg: "Design of Professional Laboratory Exercises for Effective State-of-the-Art OSINT Investigation Tools - Part 3: Maltego". Proceed. Electronic Imaging Symposium 2021 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2021 (in print)

[30] Schwarz, Klaus: "Conception and Implementation of Professional Laboratory Exercises in the Field of Open Source Intelligence (OSINT) for use in English and German Training Market for Security Authorities". Master Thesis, Technische Hochschule Brandenburg, Department of Computing and Media, April 2020

[31] Kant, Daniel; Reiner Creutzburg: 'Investigation of risks for Critical Infrastructures due to the exposure of SCADA systems and industrial controls on the Internet based on the search engine Shodan". Proceed. Electronic Imaging Symposium 2020 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2020 **https://doi.org/10.2352/ISSN.2470-1173.2020.3.MOBMU-253**

[32] Pilgermann, Michael; Thomas Bocklisch; Reiner Creutzburg: "Conception and implementation of a course for professional training and education in the field of IoT and smart home security". Proceed. Electronic Imaging Symposium 2020 (San Francisco, USA), Mobile Devices and Multimedia: Technologies, Algorithms & Applications Conference (MOBMU) 2020 **https://doi.org/10.2352/ISSN.2470-1173.2020.3.MOBMU-277**

## Author Biography

*Klaus Schwarz received his B. Sc. and M.Sc. in Computer Science from Technische Hochschule Brandenburg (Germany) in 2017 and 2020, respectively. His research interests include IoT and Smart Home Security, OSINT, Mechatronics, Sensorics, Embedded Systems, Artificial Intelligence, and Cloud Security. As a faculty member, he is developing a graduate program in Applied Mechatronic Systems focusing on Artificial Intelligence at SRH Berlin University of Applied Sciences.*

*Reiner Creutzburg is a Retired Professor for Applied Computer Science at the Technische Hochschule Brandenburg in Brandenburg, Germany. Since 2019 he is a Professor of IT Security at the SRH Berlin University of Applied Sciences, Berlin School of Technology. He is a member of the IEEE and SPIE and chairman of the Multimedia on Mobile Devices (MOBMU) Conference at the Electronic Imaging conferences since 2005. In 2019, he was elected a member of the Leibniz Society of Sciences to Berlin e.V. His research interest is focused on Cybersecurity, Digital Forensics, Open Source Intelligence (OSINT), Multimedia Signal Processing, eLearning, Parallel Memory Architectures, and Modern Digital Media and Imaging Applications.*

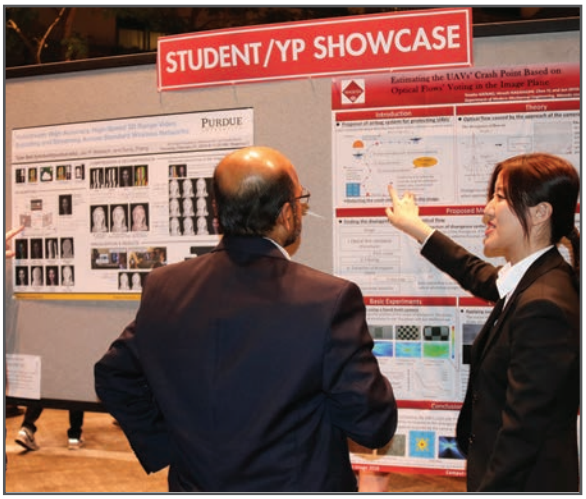
# JOIN US AT THE NEXT EI!

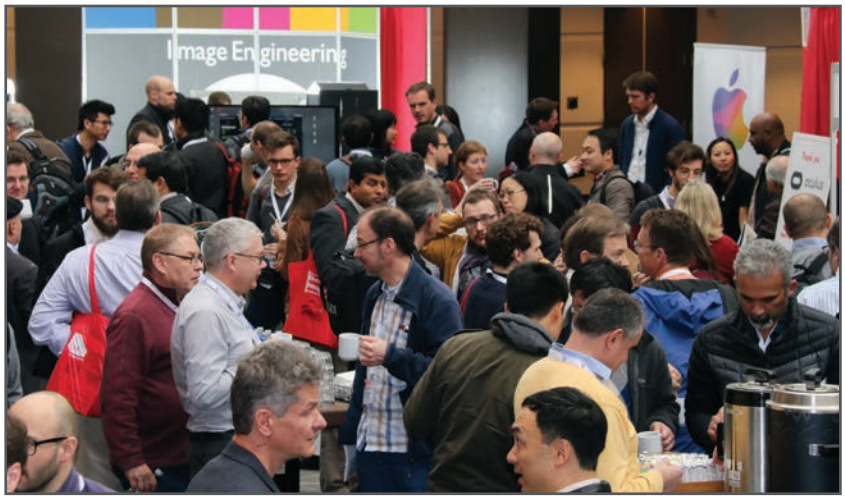
IS&T International Symposium on

# Electronic Imaging

SCIENCE AND TECHNOLOGY

*Imaging across applications . . . Where industry and academia meet!*





**• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •**
**• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •**

**www.electronicimaging.org**

IS&T
imaging.org