All that Glitters is not Gold: Uncovering Exposed Industrial Control Systems and Honeypots in the Wild

Martin Mladenov
Delft University of Technology

László Erdődi
Norwegian University of Science and Technology

Georgios Smaragdakis
Delft University of Technology

***Abstract*—Industrial control systems have enabled the digitalization and automation of industrial production and services, such as electric powerhouses, the electric grid, and water supply networks. Due to their critical role, any exposure to the public Internet makes them vulnerable to attacks that may have catastrophic implications.**

In this paper, we report that the readily available application-layer scanning on all ports opens new avenues to assess the exposure of devices that run industrial control protocols that were not possible with previously proposed active port scanning. We consider 17 widely used industrial control system protocols and develop a methodology that unveils around 150 thousand industrial control systems exposed around the globe. Our study shows that many allegedly exposed industrial control systems are honeypots that emulate industrial protocols. Our methodology infers the presence of honeypots and classifies them into three tiers based on the confidence that these act as honeypots: low-, medium-, and high-confidence. We classify them thanks to large-scale application-layer scanning on all ports and multiple independent attributes, including network information, number of open ports, and known honeypot signatures. Our results show that 15 to 25% of the exposed industrial control systems are honeypots (with two-thirds of them belonging to the medium- or high-confidence categories). Our results challenge previous reports on the prevalence and distribution of exposed industrial control systems. The developed methodology enables industry operators to assess exposed assets and aid protection teams in creating stealthier honeypots.

***Index Terms*—ICS, SCADA, Honeypots, Internet measurement.**

1. Introduction

Industrial control systems (ICS) were introduced decades ago to control and automate industrial processes. Industrial control systems have hardware and software components, including supervisory control and data acquisition (SCADA) systems and programmable logic controllers (PLCs) that receive data from sensors measuring process variables. ICS devices compare the values of the collected data with desired values or specifications and take actions based on predefined functions, e.g., tune, interrupt, or terminate a procedure.

Today, ICS devices are ubiquitous in critical infrastructures, e.g., electric grids, power producers, gas companies, production lines, smart homes, and enterprise environments. Technology giants such as Siemens, Honeywell, ABB, Mitsubishi, and IBM have invested billions of USD in the research and development of ICS products. Moreover, offering ICS solutions is a core business of many companies, e.g., Schneider Electric. Some of these companies develop their own ICS protocols, e.g., Siemens has developed S7 [1], and others contribute to and adapt industrial protocol standards, e.g., Modbus [2], BACnet [3], IEC 60870-5-104 [4].

Despite the profound impact of ICS in the digitalization of production and automation, incidents show that they are vulnerable to cyberattacks. Compromising such devices may be catastrophic. For example, Stuxnet [5], a malicious computer worm developed in the early 2000s, targeted SCADA systems and is believed to be responsible for causing substantial damage to the nuclear program of Iran. Two days before Christmas of 2015, Ukraine suffered a power outage due to a SCADA-targeted cyberattack using malware, which impacted part of Kyiv [6]. Many other high-profile attacks in power, water, and communication systems that rely on ICS devices are presented in the report prepared for the U.S. Department of Energy in 2018 [7]. Only in the first half of 2023, the US Cybersecurity and Infrastructure Security Agency (CISA) disclosed 670 vulnerabilities affecting ICS deployments [8].

Unfortunately, ICS protocols are not secure by design and often exchange traffic in plain text or with very weak authentication [9], [10]. Thus, they should not be exposed to the public Internet and should only be used within private networks. However, current best practices, e.g., firewalls and virtual private networks, are only sometimes in place [9], [11]. Understanding the current state of exposed ICS devices using network measurements is a step forward towards informing their operators of the potential risks and developing strategies to protect them from cyberattacks.

Previous studies focused on discovering devices with open ports used by some ICS protocols [12]–[16]. However, an open port is only a necessary but not sufficient condition for a host to run a protocol. With the introduction of advanced application-level scanning, which is now available through scanning companies like Censys [9], [11], [17]–[21] and Rapid7 [22], it is now possible to initiate protocol-level handshakes and report on the establishment of connections. This is an advantage for the assessment of exposed assets as hosts that passively have their ports open, e.g., non-interactive honeypots, are not

tagged as potentially exposed ICS devices.

Censys regularly scans the entire IPv4 address space on all ports [19]–[21], [23]–[25]. This allows for raw data that can be utilized to detect low- and high-interactive honeypots after enriching it with additional metadata. This is an advantage for protection teams and honeypot developers to evaluate if honeypot deployments are discovered and make them stealthier.

In this paper, we develop an algorithm and a data processing pipeline to improve the quality of detected exposed ICS devices and ICS honeypots on the public Internet. We also describe our experience in applying our algorithm on raw Censys scanning data, which we have full access to under an academic license.

Our contributions can be summarized as follows:

- This is the largest global study for Internet-facing ICS devices and honeypots in terms of discovered exposed hosts and protocols (see Section 3). We consider 17 widely used industrial control protocols, the largest ICS protocol set ever studied. We identified around 150 thousand unique IPv4 addresses that host ICS services in 175 countries by examining a complete application-layer handshake in an analytics pipeline that takes as input raw scanning data.
- Our results show that in April 2024, around 15% of addresses that run ICS protocols seem to host honeypots that mimic ICS protocols. In January 2025, the percentage of ICS honeypots increased to 25%. Our methodology and findings challenge previous ICS studies (see Section 3) which either partially considered or completely overlooked honeypots, leading to an inflated number of detected exposed ICS devices. It improves the detection accuracy of vulnerable ICS devices and makes researchers aware of current pitfalls in detection methods.
- Our study finds variations in Internet-facing ICS protocols and systems across the globe. We also show that there is a significant number of suspected ICS honeypots with an unusually large number of open ports.
- We notice that for some industrial protocols, the proportion of honeypots is very high, up to 92%. This is largely due to advanced honeypots that are able to emulate a large number of protocols, including industrial ones.
- Apart from well-studied honeypot software like Conpot [26], our results also expose large installations of advanced honeypot families located at hosting providers. We find hosts which simultaneously emulate industrial control protocols alongside others such as Bitcoin, ElasticSearch, TeamViewer, etc.
- We make our code publicly available to reproduce the results in the paper and enable future research: https://github.com/martinmladenov/ICS-Honeypots

2. Background

Industrial control systems (ICS) play a crucial role in modern society by managing and regulating the operations of various industrial processes. ICS are essential for the efficient and safe functioning of critical infrastructure such as power plants, manufacturing facilities, and transportation systems. ICS facilitate automation, enabling precise control over machinery and processes, leading to increased productivity and reduced human error. As industries continue to evolve and become more interconnected, the significance of robust and secure industrial control systems becomes increasingly apparent, ensuring not only economic efficiency but also the protection of vital infrastructure and the well-being of society at large.

ICS operate with many different protocols. The existence of various protocols in ICS is driven by the diverse and specialized nature of industrial processes and equipment. Different industries have unique requirements, necessitating protocols that are tailored to their specific needs. For instance, protocols like Modbus [2], Profibus [27], and EtherNet/IP [28] are designed to accommodate varying communication speeds, data types, and network architectures. Standardization efforts aim to foster interoperability and communication between devices from different manufacturers, contributing to a more flexible and adaptable industrial landscape. In essence, the diversity of protocols in ICS reflects the dynamic nature of industrial applications, allowing for the optimization of communication and control strategies based on the specific demands of each sector.

2.1. Industrial Control System Protocols

Because of the diverse nature of the different ICS applications, various protocols are in use. One of the most widely adopted communication protocols in industrial automation is Modbus [2], known for its simplicity and efficiency. Originally developed by Modicon in 1979, Modbus has become a standard for connecting electronic devices in various industries. Modbus is versatile and can be implemented over different communication mediums, including serial lines and Ethernet. Despite being a relatively old protocol, its continued relevance is attributed to its reliability, ease of implementation, and the extensive support it receives from a wide range of industrial devices and equipment.

Because of the different devices and variety of systems in industry automation, the development of a standard that makes equipment from many different suppliers interoperate was inevitable. As another example, IEC 60870 and IEC 61850 are both standards that play essential roles in the field of industrial automation and power systems. IEC 60870 [4], established by the International Electrotechnical Commission, defines communication protocols for telecontrol (telemetry) purposes, particularly in the context of electrical substations and power systems. It outlines the structure for information exchange between remote terminal units (RTUs) and master stations, ensuring the reliable and efficient transmission of data critical for SCADA systems.

Apart from the aforementioned protocols, many more device-specific ICS protocols are in use. The ones we study are summarized in Table 1.

2.2. Honeypots for Industrial Control Systems

Honeypots are employed to serve as decoy systems to attract and detect malicious activities, providing valuable insights into the tactics, techniques, and procedures used by potential attackers. By mimicking vulnerable systems, honeypots help their operators study and understand emerging threats, enhance threat intelligence, and fortify

TABLE 1. ICS PROTOCOLS TARGETED IN THIS PAPER

Protocol	Common ports	Information
Modbus [2]	TCP/502	Client-server communication for industrial electronic devices that can send different commands
Niagara Fox [29]	TCP/1911, TCP/4911	Automation protocol used between the Niagara software systems for managing control systems
WDBRPC [30]	TCP/17185 UDP/17185	Wind River Debug is used by VxWorks on top of RPC, it allows typical debugging functions
BACNet [3]	TCP/47808	Provides mechanisms for computerized building automation devices to exchange information
EIP [28]	TCP/44818, UDP/2222	Provides a wide-ranging, comprehensive standard to a wide variety of automation devices
IEC 60870-5-104 [4]	TCP/2404	Provides communication profile for basic telecontrol messages for power system automation
Siemens S7comm [1]	TCP/102	Used for PLC programming, exchanging data between PLCs, accessing PLC data from SCADA
ATG [31]	TCP/10001	Automatic Tank Gauging measures fuel and water levels
Codesys [32]	TCP/2455	Supports most common standard communication protocols for data exchange between controllers
Fins [33]	TCP/9600	Industrial automation control that enables seamless communication with real-time performance
OPC UA [34]	TCP/135, TCP/4840	Communication for industry 4.0 and IoT, that enables manufacturer-independent data exchange
DNP3 [35]	TCP/20000	Process automation in electric and water companies, connects data acquisition and control
PCWorx [36]	TCP/1962	Communication for inline controllers, commonly used to transmit information over long distances
ProConOS [37]	TCP/20547	High performance PLC run time engine for both embedded and PC based control applications
MMS [38]	TCP/102	Manufacturing Message Specification process real-time process data in SCADA systems
GE-SRTP [39]	TCP/18245	Transfer data from and to GE automation equipments
HART-IP [40]	TCP/5094	Transferring digital information across analog wires between smart devices and control systems

overall cybersecurity defenses. Since industrial control systems are a primary target of threat actors, the role of industrial control system honeypots is even more important.

Honeypots are commonly used both in the industry and academia. Researchers use them to characterize attacks, including threats to ICS devices [41]. Security professionals make use of honeypots to detect intruders in their networks. "Honeyfarms"—large-scale collections of honeypots deployed in multiple networks—are operated by companies such as GreyNoise [42] to collect information, which is then used to improve commercial security products [43].

One of the most well-known honeypots for ICS is Conpot [26]. Conpot is an open-source honeypot framework specifically developed for emulating ICS environments. Designed to replicate various ICS protocols and devices, Conpot helps researchers and cybersecurity professionals analyze and understand the tactics of adversaries targeting critical infrastructure. By simulating ICS components such as programmable logic controllers (PLCs) and SCADA systems, Conpot serves as an effective tool for detecting and studying cyber threats in the context of industrial networks, contributing to the enhancement of ICS security strategies and the overall resilience of critical infrastructure systems.

T-Pot [44] is a honeypot platform leveraging more than 20 different honeypot applications. It uses various honeypot technologies and services to lure attackers and gather information about their activities. It can emulate a wide range of protocols, including ones associated with ICS services.

There also exist generic low-interaction honeypots that passively listen for traffic on thousands of ports without providing application-layer responses. For example, Glutton [45] listens on all ports of the host by default. For a limited number of protocols, such as HTTP, FTP, and RDP, it provides simple application-layer responses.

There are many ICS honeypots, such as [46], [47], [48], [49], [50] for S7. There are also DNP3Pot [51] for DNP3, ShaPE [52] for GOOSE and MMS, HoneyD [53] for EIP, GasPot [14] for ATG, GridPot [54] for IEC 60870-5-104, and many more. Some can be detected and fingerprinted remotely using signatures leveraging protocol deviations, but signatures are known only for a very limited number of honeypots [55].

3. Related Work

Research on exposed ICS devices and honeypots has been an active area of study within the field of cybersecurity. Efforts to identify exposed ICS devices involve systematic scanning and monitoring of the Internet for publicly accessible ICS components. Researchers and security professionals employ various tools and methodologies to detect ICS devices that may be inadvertently exposed to the Internet, posing potential security risks. These efforts focus on discovering vulnerabilities and misconfigurations in ICS networks, raising awareness about the importance of securing these critical systems.

Exposed critical infrastructure devices in the Netherlands were analyzed in 2018 with the focus of identifying publicly available and vulnerable ICS/SCADA devices [12]. The study identified 989 potentially vulnerable devices based on well-known ICS protocols with systematic scanning and vulnerability detection. On the other hand, the study lacks detailed application-layer analysis and also the consideration of other relevant factors, e.g., the hosting organization's profile. The lack of application-layer analysis puts the accuracy of the findings in question, especially when it comes to honeypots, whose exclusion from results is only barely discussed.

Mirian et al. [56] implemented five common SCADA protocols in ZMap [9] and conducted a survey of the public IPv4 address space, finding more than 65,000 publicly accessible systems. They also used high-interaction honeypots to find and profile threat actors searching for ICS devices. Although their SCADA protocol scanning techniques were very fast, little detail about the exposed devices was analyzed.

Bitsight Security identified around 100,000 exposed ICS hosts [57]. Unfortunately, no methodology was pre-

sented on how the exposed devices were found. Therefore, it is not known how accurate their results are, and whether honeypots were considered at all.

Otorio has also published research on 108,635 exposed devices [58]. Their research was conducted as a passive search using Shodan [59]. It is unknown whether honeypots were considered.

Yaben et al. [60] identified 675,896 vulnerable IoT devices. While this study included some ICS protocols, its main focus was protocols associated with the Internet of Things. Only generic honeypots were taken into account and excluded.

Wang et al. [61] measured how long it takes for Internet-exposed ICS hosts to be updated after patches for vulnerabilities are released. They discovered 100,766 ICS devices and 1,174 honeypots. Shodan Honeyscore [62] was used to determine which are honeypots.

Wu et al. [63] proposed a method to measure the security status of Internet-facing ICS devices passively. They aggregated data from multiple public search engines and other sources. Based on open port numbers, they found 270,283 hosts across the whole study, but only 106,382 of them were identified as possible ICS based on protocol-level communication. 21,578 were classified as possible honeypots based on inconsistencies between datasets, known signatures, and the internet service provider. This work focuses mainly on vulnerability detection; the honeypot identification methodology is not discussed in detail.

Zamiri-Gourabi et al. [13] performed a large-scale Internet analysis to detect GasPot honeypots that emulate automatic tank gauges (ATG). They identified 17 GasPot honeypots among 4,853 ATG hosts. The study focuses only on GasPot honeypots and no other honeypot factors or ICS honeypot signatures were considered.

Srinivasa et al. [55] researched honeypot detection with various fingerprinting techniques. ICS specialties were not the main focus and the main services targeted were HTTP and SSH. They performed an extensive scan on the Internet that included 2.9 billion hosts and were able to identify 21,855 honeypots.

Researchers have also explored the development of realistic ICS emulation, the detection of sophisticated attacks, and the improvement of incident response strategies. Serbanescu et al. [41] have deployed a large-scale, low-interaction honeypot system on the Internet and have analysed the interactions observed during 28-day long experiments.

There is a notable challenge in the field of classifying exposed ICS devices as either honeypots or genuine operational components. Despite advancements in cybersecurity research and the development of sophisticated honeypot technologies, there has not been a widely published method that definitively distinguishes between real ICS devices and their simulated counterparts. This ambiguity poses a significant hurdle for threat actors and security researchers alike, making it difficult to ascertain the authenticity of ICS assets encountered in cyberspace. This research tries to fill this gap by using various methods to identify honeypots among the exposed devices with different methods. In addition, the existing studies do not focus specifically on ICS, do not make use of application-layer scanning, or are not as large in terms of ICS protocols or discovered hosts as our analysis. A comparison is shown in Table 2. We believe that our proposed method provides more accurate and more extensive results about the exposed ICS devices compared to the existing studies.

TABLE 2. COMPARISON OF RELATED WORK AND THIS STUDY

Study	Hosts	Honeypots	Year	Sources
[56]	65,000	0.1%	2016	Own scans
[61]	101,000	1%	2017	Shodan
[12]	989	*N/A*	2018	Shodan
[13]	5,000	0.4%	2019	Own scans
[63]	106,000	8%	2020	Shodan, Censys, FOFA, SiNan
[57]	100,000	*N/A*	2023	Unknown
[58]	109,000	*N/A*	2024	Unknown
This study	140,000 150,000	15% 25%	2024 2025	Censys, IPinfo

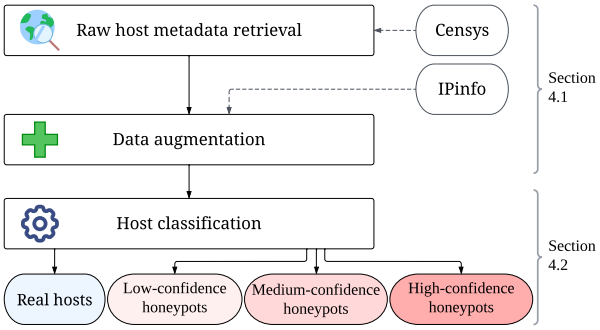

Figure 1. Main steps of our methodology.

4. Methodology

In this study, we collected the list of hosts identified to be running at least one of 17 different industrial control protocols (Table 1). We consider application-layer analysis: only hosts that successfully complete an application-layer handshake specific to an industrial control protocol are included in this list. Application-layer analysis allows for the detection of services on non-standard ports, which are common on the Internet [23]. It also enables the study to exclude hosts which are not running an industrial protocol but simply have the same ports open for another unrelated service by coincidence, or are non-interactive honeypots which do not specifically emulate industrial protocols. We found that if a simple Censys open port query is performed for the ports in Table 1, around 23 times more hosts are retrieved than with application-layer scanning. However, these additional hosts are not actually industrial control systems — they are devices running various different services which simply happen to be on the same port number as an industrial protocol or are non-interactive honeypots.

Our methodology involves (i) raw data retrieval and augmentation (Section 4.1), and (ii) data processing and honeypot identification (Section 4.2), as seen in Figure 1. We make our code available for reproducibility purposes and to enable future work in the area, see Appendix A.

4.1. Data Retrieval and Augmentation

For the raw data retrieval, we utilize a list of Internet-facing hosts running ICS protocols retrieved from Cen-

TABLE 3. COLLECTED DATA ABOUT EACH HOST AND ITS SOURCE.

Data	Source
IP address	Censys
Autonomous System Number	Censys
Autonomous System Name	Censys
Country	Censys
Domain name	Censys
Name of the IP range operator	IPinfo
Type of the IP range operator	IPinfo
Name of the AS operator	IPinfo
Type of the AS operator	IPinfo
List of open ports and identified protocols	Censys
Metadata about identified services	Censys

sys [11], [17]. Censys is a search engine which indexes the open ports of IPv4 hosts exposed to the public Internet. Censys continuously performs application-level scanning on the whole IPv4 address space on all 65,536 ports [19]–[21], [23]–[25]. Censys provides information about the location, the autonomous system number, a list of open ports, and some metadata collected from each service [17]. For this study, we had access to the complete raw scanning data of Censys under an academic license.

This approach helps reduce the scanning load that would have been needed to perform Internet-scale application-level scanning for all IPv4 space and all ports that Censys already performs. Even if we performed such a large-scale scan, Censys is expected to have better coverage than our independent scanning due to the deployment of many and distributed vantage points to improve visibility [64] and optimized scanning operations [23], [24], [65]. Moreover, its coverage is much better than other datasets, e.g., Shodan [59] and Rapid7 [22], which scan only for a set of ports and less frequently than Censys [25]. By utilizing Censys data with appropriate date and protocol information, we also allow other researchers to reproduce and compare against our results as they can get access to the same data via an academic or commercial license.

We filter the hosts in the dataset to retrieve the ones identified by Censys to be running at least one of the 17 industrial control protocols shown in Table 1. Afterwards, we enrich the collected raw data with metadata from IPinfo [66]. IPinfo provides information about the organizations that operate autonomous systems and IP address ranges, including the type of organization (business, ISP, hosting, or education). IPinfo is a commercial service, but we requested and were granted full access to the complete database under an academic license. We acknowledge the limitation that geolocation may not always be accurate, especially for network equipment devices [67], [68]. However, it is typically accurate at the country level [69], [70].

In this study, we collected data over a period of one year between January 2024 and January 2025: on January 9th, 2024, March 18th, 2024, April 29th, 2024, August 6th, 2024, October 23rd, 2024, and January 28th, 2025. The data we collect about each host is summarized in Table 3.

We perform our main analysis using the results of the April 2024 snapshot and compare the results with the latest one (January 2025). In April 2024, we found around 140,000 unique hosts in 175 countries, running at least one of 17 industrial protocols.

It is important to note that our methodology may take any raw data source as input, e.g., data from independent scanning, Rapid7, or Shodan; thus, it is not limited only to Censys data. A limitation of relying on Censys data is that Censys only very recently introduced IPv6 scanning support; thus, IPv6 application-level scanning data is limited at the time of writing this paper. We leave as part of our future work to analyze IPv6 data when it becomes available and IPv6 hit lists are more complete and accurate [71]. Our pipeline is fully automated and allows for automatic data collection and analysis, thus, it can easily be applied to IPv6 data.

4.2. Identifying ICS and Honeypots

Honeypot software is used to emulate ICS protocols in order to analyze behavior of threat actors, as discussed in Section 2.2. However, such honeypots can significantly affect the results of Internet-wide studies on industrial control systems and inflate the number of exposed ICS devices. Therefore, part of our methodology attempts to identify honeypots and exclude them from the general analysis. Based on the characteristics exhibited by a particular host, our algorithm classifies it as a potentially real ICS device or a honeypot with high, medium, or low confidence.

Signatures. In some cases, honeypot software can be fingerprinted using signatures. Such signatures can, for example, test for inaccuracies in the emulated application-layer protocol, as in the case of ATG emulation in *GasPot* [14]. Other types of signatures look for a response containing values that are part of the default configuration of known honeypot software, which are impossible to be present in that of a real device. For instance, some honeypot software for the Siemens S7 communication protocol can be fingerprinted based on this, such as *Conpot* [26] and the *Snap7 framework* [72] (see Table 4). If at least one exposed protocol of a host matches a known signature of honeypot software, the host is classified as a *honeypot* with *high confidence*.

As signatures, we used improper protocol emulation [73] as well as the presence of default honeypot configuration values in the response such as a preset serial number, a nonexistent model number, or particular content of a user-specifiable text field from their configurations [15], [16], [74]. The signatures we used in our analysis are shown in Table 4. By definition, signatures are very specific to particular honeypot software and protocols. Signatures are thus only known for some honeypot software [55]. As new signatures are discovered in the future, the list of signatures can easily be extended. We avoided using signatures that require data beyond what is provided by Censys; this improves the scalability of our methodology by enabling it to work with passive Censys queries without the need for active probing.

Network type. Industrial control systems are expected to be on industrial networks such as factories, power grids, or other commercial, residential, or government networks. It would be highly unusual for such systems to be at datacenters. Therefore, we consider hosts located at hosting providers as *honeypots* with *medium confidence*. Honeypots are known to be present on cloud networks [76], [77].

TABLE 4. SIGNATURES FOR FINGERPRINTING ATG AND S7 HONEYPOTS. [15], [16], [73], [75]

Software	Protocol	Signature
Conpot	ATG	Banner contains "\n\n\n\n"
Conpot	ATG	Timestamp format: "MM/DD/YYYY HH:MM"
Conpot	S7	Plant ID: "Mouser Factory"
Conpot	S7	Serial number: "88111222"
GasPot	ATG	Banner contains "\n\n\n\n"
GasPot	ATG	Timestamp format: "MM/DD/YYYY HH:MM"
Snap7	S7	Memory card serial number: "MMC 267FF11F"
Snap7	S7	Serial number: "S C-C2UR28922012"
Snap7	S7	System name: "SAAP7-SERVER"

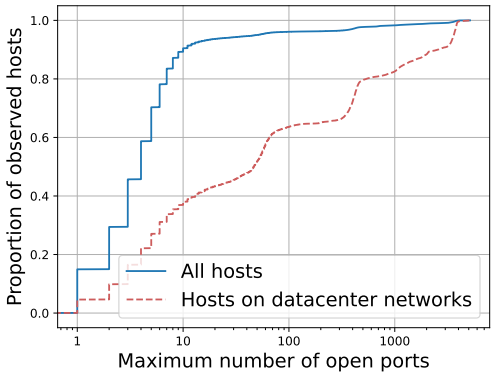

Figure 2. CDFs of the open port count of all identified hosts and hosts on datacenter networks (April 2024).

Previous work on honeypot identification has made use of this fact [55], [78].

Furthermore, it is known that researchers deploy honeypots on academic networks for research purposes [79]. Despite this, as the presence of actual ICS devices on university campuses is common, this alone is not a strong honeypot indicator. Our algorithm marks hosts on academic networks as *hosts of interest*. This does not yet assign a classification label but is used later in the algorithm.

Open ports. It is unusual for an ICS host to have many open ports; in general, without taking device specifics into account, hosts running more than 100 services are considered suspicious [80]. ICS devices are specialized for industrial settings and operate using industrial protocols; they are expected to host fewer services than general-purpose servers. Thus, a lower threshold is more appropriate. In Figure 2, we plotted the distribution of the number of open ports of the hosts in our dataset. We discovered that 89.2% of hosts have at most 10 open ports, and 94.2% have at most 30. In all types of networks except hosting providers, the vast majority of exposed ICS hosts exceed neither of these thresholds. In hosting providers, we find the opposite — the majority of hosts (62.0%) have more than 10 open ports. Taking this into account, if the number of open ports exceeds the higher threshold ($\tau > 30$), the host is considered as a *honeypot* with *medium confidence*. If a host exceeds only the lower threshold ($\tau > 10$), it is marked as a *host of interest* and a classification label is not yet assigned.

Some ICS devices may operate multiple protocols, e.g., for cross-vendor compatibility. However, the presence of completely unrelated ICS protocols on the same device, such as ones used in different types of infrastructure, is very unusual. This behavior could be used as a honeypot indication; however, we found that the vast majority of devices in our study that host multiple ICS services are already identified as honeypots by other methods. We concluded that such a heuristic would have little impact on the results of our study. We discuss unusual protocol pairs in Section 6.6.

Hosts of interest. As mentioned in this section, the algorithm marks some hosts as hosts of interest if they exhibit some behaviour consistent with being honeypots, but to an insufficient extent to confidently classify them as such. If a host is marked as a host of interest by multiple independent metrics, it is classified as a *honeypot* with *medium confidence*. If it matches only one such characteristic, it is classified as a *low-confidence honeypot*. In our analysis, we took a conservative approach and only considered high-/medium-confidence honeypots as such.

Real hosts. Finally, all hosts that do not show any indication of being a honeypot and have not yet been assigned a classification label are considered *potentially real*. Note that this does not guarantee that a host is certainly not a honeypot — it is impossible for any classification methodology to exhaustively identify all honeypots.

Honeypots can significantly affect the results of Internet-wide studies on industrial control systems. It is important to identify as many honeypots as possible and exclude them from general results to avoid inflation of the number of exposed ICS devices and misleading observations.

In this study, we found a large number of ICS honeypots — in April 2024, we found 20,342 unique IPs (15%) hosting suspected honeypots, out of which 13,591 with medium or high confidence. They were located in 2,298 autonomous systems across 121 countries. Details are shown in Tables 5 and 6 in Section 5. In January 2025, the percentage of honeypots increased to around 25%, see Table 7 in Appendix B. The magnitude of this number illustrates the importance of taking honeypots into account when performing large-scale Internet studies.

In our analysis, we used the signatures for ATG in Table 4 only for data collection in April 2024 and later. This allowed us to detect a higher number of high-confidence ATG honeypots. However, this increase is moderate and does not fundamentally change any of the results. As shown in Section 4.3, most of these honeypots had already been classified as such via other heuristics.

4.3. Validation

We made use of several techniques to validate our methodology. This includes setting up our own hosts as ground truth (Section 4.3.1), contacting operators of devices (Section 4.3.2), comparison with Censys labels (Section 4.3.3), and comparing heuristics against known signatures (Sections 4.3.4 and 4.3.5).

4.3.1. Self-Operated Honeypots. As ground truth, we set up two honeypots ourselves in order to see whether they would be detected by our methodology. Namely, we configured a Conpot instance [26] on a server in Norway and a T-Pot instance [44] in Hungary. We used their respective default configurations. We found that they were detected by Censys within days of being placed online,

and our classification algorithm successfully identified both as honeypots.

Conpot. The Conpot instance had 10 open ports, among which were the industrial protocols S7, Modbus, and BACnet. It was fingerprinted as a high-confidence honeypot based on its known signature; namely, the default values used for the S7 communication protocol.

T-Pot. The T-Pot instance had 64,527 open ports. It exposed 16 open ports in the range 1-1023 and all ports in the range 1024-65535. T-Pot uses a Conpot instance as one of its components, albeit in a non-standard configuration, enabling it to avoid some signatures [44]. Nonetheless, it was classified as a high-confidence honeypot based on inaccurate emulation of the ATG protocol.

4.3.2. Responsible Disclosure. We made contact with several organizations on whose networks we identified exposed devices. We were able to confirm the accuracy of our classification with some of them. We conducted the disclosure by sending emails to security-related addresses (if provided on the organization's website) or to general contact addresses. In our emails, we included a list of IP addresses and protocols and provided general information about our research as well as our contact information.

We found one exposed device on the network of one of the largest energy providers in Northern Europe, which we suspected to be a real ICS device. We contacted the company and received confirmation that the host was indeed not a honeypot and that it was accidentally exposed. The responsible national CERT for critical infrastructure was involved and action was taken to resolve the issue.

We contacted a research network operator in the Netherlands regarding 4 hosts suspected as honeypots. We received confirmation from the operator that the devices were set up by cybersecurity researchers and that the exposure of those services was intentional, so it is highly likely that those hosts are indeed honeypots.

We contacted an ISP in the Netherlands regarding 246 exposed industrial control systems which we classified as real. The ISP replied that the security team would consider informing the customers.

We also contacted 4 industrial plants in Germany regarding exposed controllers. Unfortunately, 90 days after our notifications, we had not received responses from any of them, and the devices remained online.

Due to the large number of exposed hosts, notifying each operator individually would be infeasible [81]. We are currently in the process of notifying the responsible national CERTs of the exposure by automatically preparing lists of exposed IPs in each country and sending emails.

4.3.3. Comparison with Censys Labels. Censys automatically assigns labels to suspicious hosts based on heuristics. Among those are the labels "truncated" (hosts with more than 100 open ports), "tarpit" (hosts trying to trigger rules in security software), and "honeypot" [73], [80], [82].

We compared our results with the labels assigned by Censys and found that 68.7% of the hosts we identified as honeypots have at least one of these labels. In contrast, among the devices we classified as real, there were only 10 such hosts. Therefore, the hosts marked as suspicious by Censys' heuristics are almost entirely a subset of the ones our methodology classifies as honeypots.

We note that Censys scans more ports and hosts than Shodan [25]; thus, this tagging has better coverage than Shodan Honeyscore [62]. During our study, Shodan Honeyscore returned error messages for the honeypots we operated as ground truth and other hosts for which we had strong indications that they were honeypots, e.g., due to the high number of open ports. Therefore, we decided not to consider Shodan Honeyscore service feedback in our study.

4.3.4. Case Studies. As described in Section 4.2, some honeypots can be identified via fingerprinting based on known signatures of honeypot software. Such signatures are available for ATG and S7 [15], [16], [73], [75]. We used these high-confidence signatures to evaluate the sensitivity of our other heuristics that aim to identify honeypots by network type and number of open ports.

Case study: ATG. We discovered 1,612 high-confidence honeypots matching signatures that look for deviations from the ATG protocol. However, even if signature-based fingerprinting had not been performed, we found that 99.8% of these certain honeypots would also have been classified as honeypots by one of the heuristics.

Case study: S7. We identified 115 Siemens S7 communication protocol hosts as high-confidence honeypots matching signatures. Without the use of signatures and instead considering only the network type and port count heuristics, we found that 91.3% of them would still be classified as honeypots.

The ATG and S7 case studies show that for such honeypots, detection based on network information and the number of open ports has high sensitivity and is largely sufficient. This shows that even if signatures are unavailable or become outdated, these hosts can still be detected.

4.3.5. Reply Signatures. In [55], possible reply signatures for S7 and Modbus honeypots are described. In contrast to the signatures used in our methodology, checking for these signatures requires active probing by sending a specific packet to each host and looking for deviations from the protocol. To the best of our knowledge, reply signatures for other industrial protocols are not publicly available.

Modbus. [55] describes a signature for identifying Modbus honeypots. After sending a packet with specific content to the host, a real device is expected to return a response. If the host disconnects, it is considered a honeypot. In our testing, our own Conpot honeypot matched this signature.

Since we relied on passive measurement via Censys data, this signature detection was not there, as it requires active probing with a specific message. According to [55], this signature applies only to Conpot honeypots. Thus, our algorithm cannot fingerprint Conpot honeypots based on the Modbus service alone. Overall, around 1% of the devices classified as real by our algorithm match the Conpot reply signature, but the large majority of them are concentrated in two ISPs in Turkey, and this could be due to specific configuration. It is possible that many of the devices that do not match this signature are also

TABLE 5. PROTOCOLS AND AGGREGATED DATA. FOR EACH PROTOCOL, THE NUMBER OF DETECTED POTENTIALLY REAL DEVICES (REAL) AND THE NUMBER OF SUSPECTED LOW-, MEDIUM-, AND HIGH-CONFIDENCE HONEYPOTS (HP) ARE SHOWN. THE NUMBERS OF AUTONOMOUS SYSTEMS AND COUNTRIES WITH AT LEAST ONE POTENTIAL REAL DEVICE OR AT LEAST ONE SUSPECTED HONEYPOT ARE ALSO SUMMARIZED.

	January 9th, 2024						March 18th, 2024						April 29th, 2024					
	Hosts		ASes		Countries		Hosts		ASes		Countries		Hosts		ASes		Countries	
Protocol	Real	HP	Real	HP	Real	HP	Real	HP	Real	HP	Real	HP	Real	HP	Real	HP	Real	HP
MODBUS	43,306	3,527	2,095	778	133	77	48,522	3,732	2,154	768	131	74	49,040	4,511	2,015	871	134	75
FOX	20,464	2,806	1,182	573	71	49	21,358	2,854	1,164	573	72	50	20,646	3,140	1,083	642	72	52
BACNET	12,291	1,566	1,119	435	96	60	12,662	1,578	1,122	462	98	66	11,449	1,638	1,003	485	88	66
EIP	10,795	825	617	205	82	58	11,353	1,085	629	215	83	62	12,455	1,145	614	235	85	71
IEC60870_5_104	6,839	2,196	222	383	58	85	7,478	1,509	236	416	51	86	7,673	2,007	221	422	52	87
S7	7,848	926	658	228	81	51	8,836	738	682	214	81	53	8,792	920	644	248	83	58
ATG	5,344	1,872	452	358	39	82	5,427	1,595	462	377	39	86	5,409	2,156	466	408	38	90
WDBRPC	2,868	3,476	415	206	91	65	2,855	3,744	398	210	87	64	4,101	6,348	535	497	93	81
CODESYS	3,026	234	340	132	59	26	3,130	231	349	112	53	26	2,995	245	314	124	53	29
FINS	2,649	187	279	83	58	35	2,220	207	293	83	60	33	2,041	271	273	94	61	37
OPC_UA	1,788	1,032	410	247	72	62	2,067	1,287	432	264	73	68	2,111	1,444	411	293	69	67
DNP3	526	355	77	52	32	42	572	507	83	56	27	43	585	535	81	55	29	44
PCWORX	476	11	59	7	20	7	560	23	62	7	18	6	620	28	65	8	18	7
PRO_CON_OS	35	304	10	27	6	38	35	460	11	26	6	38	41	476	14	30	7	41
MMS	44	25	14	9	11	7	48	24	15	10	9	7	58	30	21	14	13	9
GE_SRTP	53	10	32	8	18	6	55	6	33	6	17	5	54	9	33	6	16	5
HART	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
Total	110,217	15,685	4,017	1,894	169	114	118,469	15,587	4,105	1,935	164	114	119,534	20,342	3,924	2,298	170	121

TABLE 6. NUMBER OF HOSTS IN EACH CLASSIFICATION CATEGORY AND THE NUMBER OF ASES AND COUNTRIES WITH AT LEAST ONE HOST IN A GIVEN CATEGORY (APRIL 29TH, 2024).

	Hosts				Autonomous Systems				Countries			
	Real	Honeypots			Real	Honeypots			Real	Honeypots		
Protocol		*Low*	*Medium*	*High*		*Low*	*Medium*	*High*		*Low*	*Medium*	*High*
MODBUS	49,040	2,527	1,981	3	2,015	493	442	3	134	69	61	3
FOX	20,646	1,794	1,346	0	1,083	405	273	0	72	45	42	0
EIP	12,455	401	663	81	614	106	105	42	85	35	52	33
BACNET	11,449	922	716	0	1,003	297	204	0	88	53	47	0
WDBRPC	4,101	509	5,839	0	535	129	394	0	93	38	72	0
S7	8,792	211	584	125	644	95	111	59	83	28	45	37
IEC60870_5_104	7,673	122	400	1,485	221	45	87	336	52	20	46	86
ATG	5,409	51	493	1,612	466	39	33	354	38	5	39	86
OPC_UA	2,111	465	978	1	411	149	153	1	69	50	57	1
CODESYS	2,995	161	84	0	314	70	56	0	53	21	21	0
FINS	2,041	81	189	1	273	48	49	1	61	23	30	1
DNP3	585	26	508	1	81	12	43	1	29	10	41	1
PCWORX	620	27	1	0	65	7	1	0	18	6	1	0
PRO_CON_OS	41	0	476	0	14	0	30	0	7	0	41	0
MMS	58	5	25	0	21	5	9	0	13	4	6	0
GE_SRTP	54	7	2	0	33	4	2	0	16	3	2	0
HART	1	0	0	0	1	0	0	0	1	0	0	0
Total	119,534	6,751	11,878	1,713	3,924	1,123	1,170	388	170	93	94	91

honeypots, but not Conpot honeypots. Indeed, for 1,186 out of 2,501 honeypots that our algorithm detected and do not match the Modbus reply signature, the number of open ports is very high (more than 10). The large majority (30,369; 93%) of Modbus hosts that replied to our active requests and identified as ICS do not match the Modbus reply signature.

S7. This signature relies on a honeypot returning an unexpected response when certain protocol specifications are not followed during interactions: [55] describes the process of sending only the third packet in the handshake sequence used by Nmap [83] instead of the whole sequence. Responding to this malformed handshake would be a protocol deviation, indicating that the host is a honeypot. In our testing, we observed that a real device would indeed instead disconnect.

In our validation, we connected to 6,977 S7 hosts, out of which our algorithm classified 699 as suspected honeypots (85 based on a signature and 614 based on network information), and 6,278 as real. In [55], it is mentioned that this signature can be used to detect Conpot honeypots. However, out of the 70 Conpot honeypots fingerprinted by the technique described in Section 4.2, only 2 matched this signature. The rest disconnected, as would be expected by a real device, including our own Conpot honeypot. However, the signature matched on all 15 of the Snap7 server framework honeypots fingerprinted by our methodology. It is possible that the Conpot signature has changed in newer versions released after the methodology of [55] was made. Our passive methodology was able to identify 157 out of 183 (85.8%) honeypots fingerprinted by the active measurement technique. Again, the large majority, i.e., 6,252 (89.3%) of the S7 ICS devices that we scanned do not match the reply signature and are also not suspected as honeypots by our methodology.

4.4. Ethical Considerations

For our study, we do not perform active measurements or exchange traffic with the discovered hosts to collect

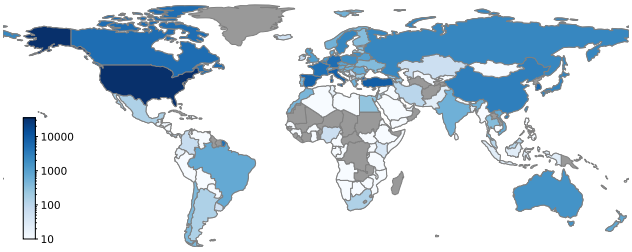

Figure 3. Geographical distribution of hosts classified as real (April 2024).

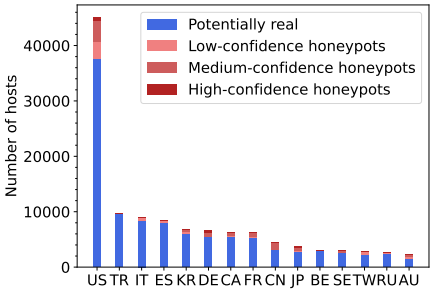

Figure 4. Top 15 countries with the highest number of exposed hosts (April 2024).

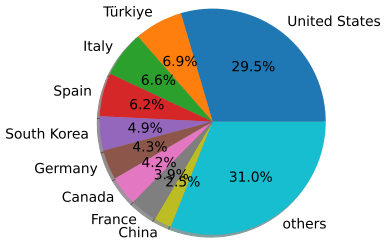

Figure 5. Distribution of real ICS devices per country (April 2024).

raw data. The only exceptions to that are selected active measurements for validation and honeypot investigation. For this, we connected to a subset of all hosts and retrieved basic details to validate the results (as described in Section 4.3). We also connected to some hosts to investigate certain unusual honeypots (in Section 6.7). When connecting to hosts, we followed the Recommended Practices described by Durumeric et al. [9], namely, coordination with local network administrators, usage of a server with a dedicated IP address for connection initiation, a descriptive DNS entry, and a website with an explanation of our research hosted on the server along with contact details. When we identify a potential vulnerability, we do not exploit it in any way and report it to the competent national Computer Emergency Response Team (CERT).

5. Demographics of Exposed ICS Devices

In this section, we study if there are noticeable differences across regions, countries, and protocols. This is important to identify locations where there are more honeypots than the global baseline or dominance of particular ICS protocols and honeypots. We first consider the overall global picture (Sections 5.1 and 5.2), then we look into individual countries (Section 5.3), then into autonomous systems (Section 5.4), and finally, we dive into some interesting hosts or groups of hosts (Section 5.5). In our analysis, we focus on devices from the April 2024 dataset classified as real, unless otherwise specified.

5.1. Empirical Observations

In our study, we collected data over a period of one year between January 2024 and January 2025. We found that the absolute numbers of real devices and the proportion of suspected honeypots continuously increase over time across our scan snapshots, as can be seen in Table 5 (January 2024–April 2024) and in Appendix B (August 2024–January 2025).

5.1.1. Churn. We noticed that the sets of IP addresses observed during the study changed significantly. Only 68% of the real devices observed in January 2024 were also seen in April 2024. Between the two snapshots, 33,869 real hosts disappeared and 45,106 new ones appeared. The vast majority of these IPs belong to Internet service providers. We manually investigated the reverse DNS records of these hosts and most of them appeared to be hosted in residential and commercial networks that use DHCP. The number of hosts within each AS between these two snapshots also does not significantly deviate. This symmetry is an indication that the hosts are largely the same, but their IP addresses changed in that time due to DHCP. Overall, we found no significant qualitative changes in the results within that period in terms of numbers and classification proportions of devices within ASes, countries, or protocols.

5.1.2. Global Overview. Overall, considering all 17 protocols part of this study, we found 139,876 hosts in 5,497 autonomous systems, located in 175 countries. They were split into two categories: (i) 119,534 potentially real ICS devices in 3,924 ASes and 170 countries, and (ii) 20,342 suspected honeypots in 2,298 ASes and 121 countries, as seen in Tables 5 and 6. Modbus is the most popular exposed protocol globally, used by 38.3% of ICS devices classified as real. The next most popular is Niagara Fox, used by 16.1%. Third is EtherNet/IP with 9.7%, and BACnet follows closely with 8.9%.

5.2. Global Device Distribution

The number of devices identified as potentially real in each country is shown in Figure 3. The total number of exposed hosts and their classifications in the top countries by the number of hosts is shown in Figure 4. The US stands out with more than 45,000 exposed devices. All other countries have fewer than 10,000 exposed devices. It is worth noting that the number of suspected honeypots varies in the top countries and does not seem to be proportional to the total number of ICS per country. We go into more detail on honeypots in Section 6. The distribution of potentially real devices is also shown in Figure 5. About a third of all exposed hosts are in the US, around a third in the following 7 countries, and another one-third in the rest of the world.

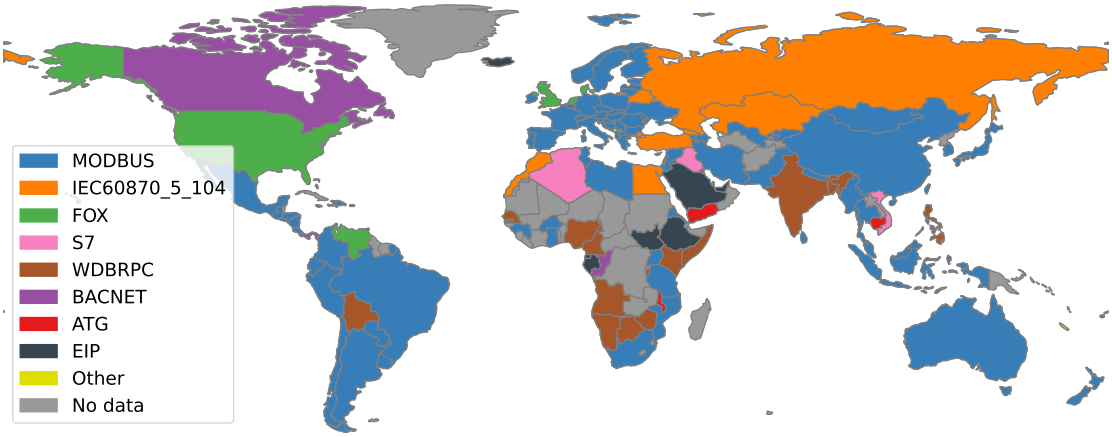

Figure 6. Most popular exposed industrial protocols in real devices per country (April 2024).

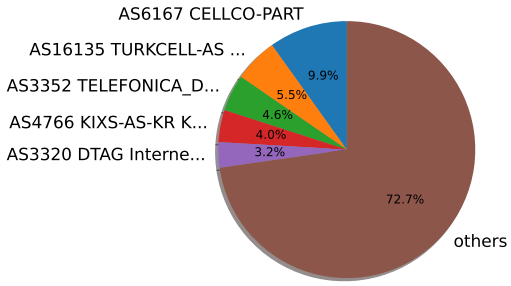

Figure 7. Distribution of real ICS devices per AS (April 2024).

5.3. Protocol Popularity per Country

As discussed earlier in this section, Modbus is the most popular exposed ICS protocol globally. However, we observe geographical differences in terms of protocol popularity. Figure 6 shows the most popular exposed ICS protocols per country among real devices. In most of Europe, South America, Australia, and China, Modbus is the most popular exposed protocol. In the US, this is Niagara Fox. In Russia, Belarus, Turkey, and Kazakhstan, this is IEC 60870-5-104. BACnet is the leading protocol in Canada. Thus, we notice fragmentation of the leading exposed protocol in different geographic regions.

We also see significant differences in the distribution of observed industrial protocols within individual countries. Niagara Fox and BACnet are popular mainly in the US and Canada. In Turkey, while IEC 60850-5-104 is the most popular protocol, Modbus is almost as popular. In Belgium and South Korea, Modbus is exposed by around three-quarters of all real ICS devices, a much higher proportion than in other countries. The full protocol distributions in the top 10 countries can be seen in Figure 16 in Appendix C.

5.4. Popularity per AS

We now turn our attention to autonomous systems hosting ICS. Figure 7 shows the top ASes with the most exposed devices. The top 5 forms about a quarter of all discovered devices, and they all belong to ISPs. Out of all autonomous systems with ICS classified as real, we found 3,535 (93.5%) are ISPs, 243 (6.4%) belong to businesses, and 1 is unclassified (based on IPinfo data). In terms of IPs, we observed that 98.3% are in ASes that belong to ISPs, 1.7% are in ones that belong to businesses, and 5 IPs are unclassified.

We observe that in a large majority of cases, each AS hosts only a few industrial protocols among all of its hosts. Moreover, examining the number of protocols seen in individual devices, we see that the vast majority of hosts (94%) only run one. Running many industrial protocols would be unusual for a real ICS device.

5.5. Oddballs

In this subsection, we elaborate on some interesting findings. We found 388 devices that share IP addresses with email servers, based on a DNS lookup. They appear to be real controllers behind NAT (Network Address Translation). We randomly picked some such devices for manual investigation. Hosted on the same IP addresses, we indeed found the email and web servers of industrial plants. Around half of the discovered devices were in the US and one-eighth in Canada. The UK, Russia, Germany, and Belgium each had around a dozen such hosts. In terms of protocols, 177 of them hosted Niagara Fox, 102 had BACnet, 57 - Modbus, and the rest had other protocols.

China has a surprisingly small number of exposed hosts running industrial protocols with regards to its size and population — we observed only 4,402 devices. We are not aware of any previous work showing that the Great Firewall of China blocks entire protocols. In our data, we also do not see evidence suggesting this to be the case for industrial protocols — we do observe some hosts in China for all major protocols we target. This suggests that firewall-level protocol blocking is indeed unlikely to affect those protocols. We believe that this is likely caused by the scarcity of IPv4 addresses in the country and the use of carrier-grade NAT [84].

6. Deep Dive into ICS Honeypots

In this section, we turn our attention to the hosts suspected as honeypots based on our methodology. We first consider the global baseline (Section 6.1). Next, we

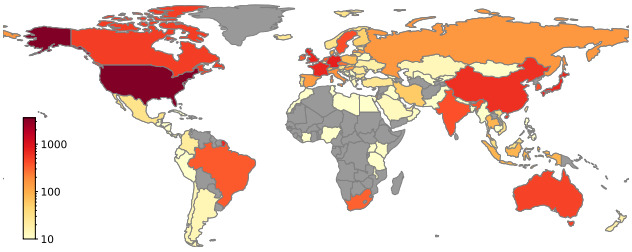

Figure 8. Number of suspected high- and medium-confidence honeypots per country (April 2024). Countries with no such hosts are shown in grey.

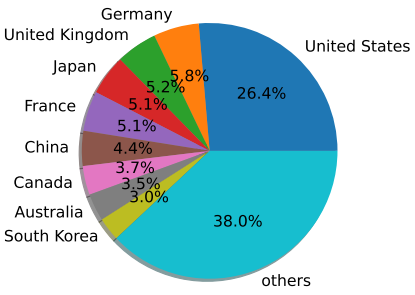

Figure 9. Distribution of suspected high- and medium-confidence honeypots per country (April 2024).

break down the results by country (Sections 6.2 and 6.3). Then, we consider autonomous systems (Section 6.4), and finally, we consider interesting findings related to groups or types of honeypots (Sections 6.5, 6.6, and 6.7). In our analysis, we take a conservative approach with regard to honeypot classification; we only consider honeypots that were classified with high or medium confidence, unless otherwise noted.

6.1. Global Honeypot Distribution

As shown in Table 6, we found 20,342 unique hosts suspected as honeypots of any confidence level. We observed 1,713 high-confidence honeypots (identified via a signature), 11,878 medium-confidence ones, and 6,751 were classified with low confidence. The most popular industrial protocol used by suspected medium- and high-confidence honeypots is WDBRPC, hosted by 5,839 unique IPs, followed by Modbus, observed on 1,984 IPs. An interesting observation is that some protocols appear to have a higher honeypot proportion than the baseline. Apart from WDBRPC (60.8%), we also observe high honeypot proportions in OPC UA (40.6%), DNP3 (47.8%), and ProConOS (92.0%).

6.2. Popularity of Honeypots per Country

Figure 8 shows the geographical distribution of suspected honeypots. The US is home to 4,442 suspected honeypots, around a quarter of the total number of honeypots found worldwide. Turkey has surprisingly few suspected honeypots in comparison to the number of real devices - only 32 of the 9,709 hosts in Turkey are suspected as medium- or high-confidence honeypots and 150 are considered low-confidence ones.

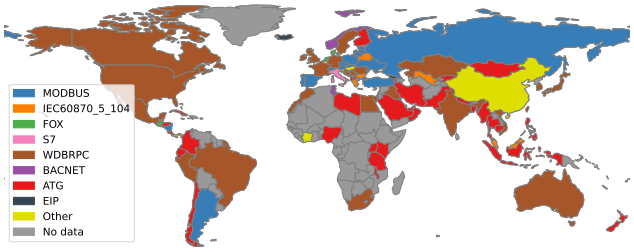

Figure 10. Most popular industrial protocols per country used by suspected high- and medium-confidence honeypots (April 2024).

In comparison to devices classified as real, the set of top countries differs. Figure 9 shows that while the US once again has the top place with about a quarter of the detected hosts, Turkey, which was in second place with about 6.9% of all global devices classified as real, is not in the top 10 countries with the most suspected honeypots. On the other hand, the UK, which is not in the top 10 with the most potentially real devices, has the third-largest number of suspected honeypots.

In honeypots, the distributions of exposed protocols per country were much closer to the global baseline than in real devices. There is significantly less geographical variance in honeypot protocols, possibly due to the availability of honeypot software. The full distributions in the top 10 countries are shown in Figure 17 in Appendix C.

Figure 10 shows the most popular industrial protocol used by suspected honeypots in each country. We notice that in most countries, WDBRPC and Modbus are the most popular, followed by ATG, Fox, and BACnet. In China, the most popular exposed honeypot protocol is OPC UA, which is not nearly as common in other large countries. Nonetheless, it is closely followed by Modbus.

6.3. Proportion of Honeypots per Country

We would like to put the number of discovered honeypots per country in the context of the total number of exposed ICS hosts there. Overall, across all countries in our study, in April 2024 around 15% of all identified hosts are suspected to be honeypots (and around 25% in January 2025). However, we observe geographical deviations from this baseline. Figure 11 shows the proportions of hosts in each country which have been classified as honeypots. There are big differences between countries: in much of Europe and North America, only a small proportion of all exposed hosts are identified as honeypots, while in Ireland, Indonesia, Saudi Arabia, India, Brazil, the United Kingdom, China, and others, this proportion is much higher. While it is difficult to extract a rule as to why this is the case, it is apparent that not all countries have the same demographics of exposed industrial control systems and associated honeypots. Due to the large honeypot proportions observed in some countries or protocols, it is of utmost importance to exclude honeypots from the overall results of any study. If honeypots are not taken into account, results would be highly biased.

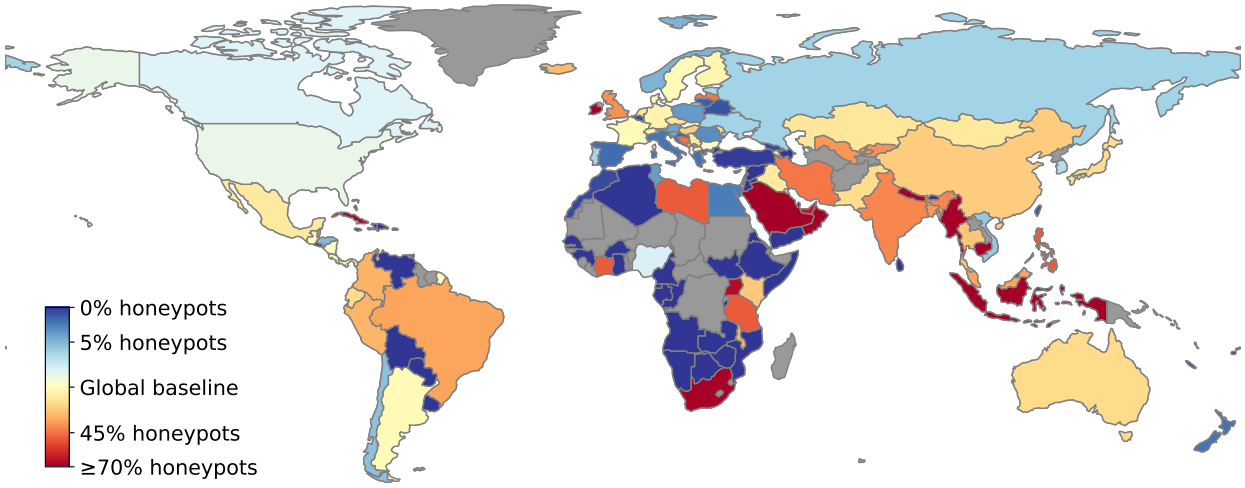

Figure 11. Proportions of exposed hosts suspected as high- and medium-confidence honeypots, per country (April 2024). Countries without any detected hosts are in grey.

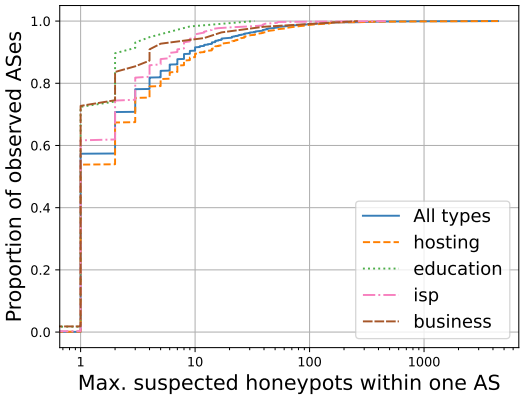

Figure 12. CDFs of the number of suspected honeypots within ASes, per type, excl. ASes with none (April 2024).

6.4. Honeypot Popularity per AS

In this section, we turn our attention to the autonomous systems hosting suspected honeypots. In Figure 12, we show the CDF of the number of honeypots hosted by each AS type, provided by the IPinfo dataset. About 60% of all ASes have only one honeypot, with the exception of the "education" and "business" (enterprise) types, where this is the case for around 73% of ASes. The CDFs of all AS types have long tails, with some ASes hosting hundreds of honeypots.

Then, we focus on the ASes with the largest number of honeypots, as shown in Figure 13. Almost 40% of all suspected honeypots are on two autonomous systems owned by Amazon — AS16509 and AS14618. In the top 5, we see two other hosting providers — DigitalOcean (AS14061) and Microsoft Azure (AS8075). However, we also notice Comcast (AS7922), a large US internet service provider. We manually analyzed some hosts in this AS. They are very likely to indeed be honeypots, based on the large variety of unrelated exposed services (WDBRPC, Chargen, FTP, RDP, Ubiquiti, DNS, NTP, Mumble, L2TP, DB2, SNMP, among others), many of which are not in any way related to ICS.

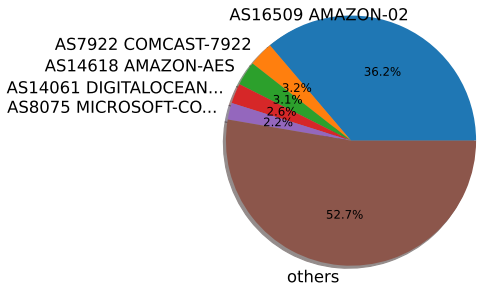

Figure 13. Distribution of global honeypots per AS (April 2024).

6.5. Characteristics of Honeypots

We look into some characteristics which we found to be common between honeypots:

Business type. Examining the distribution of suspected honeypots by business type, we see that about three-quarters of the found IP addresses are owned by hosting providers, 13% by ISPs, 9% are enterprise networks, and 2% are academic networks.

Protocol-specific honeypot characteristics. Our analysis shows that a significant number of suspected honeypots have an unusually large number of open ports. In Figure 14, we plot the distribution of the number of open ports seen in suspected honeypots, per ICS protocol. Although only 10.8% of all hosts have more than 10 open ports, this proportion is much higher among suspected honeypots: 65.6% of them have more than 10 open ports, 26.8% have more than 100, and 11.8% have more than 1,000. We also notice striking differences between protocols: for example, less than half of the detected ICS honeypots that emulate Fox or BACnet have more than ten open ports. This is in contrast with honeypots emulating WDBRPC, ATG, or IEC 60870-5-104, where the majority have hundreds or even thousands of open ports.

6.6. Multi-protocol Honeypots

We also study suspected honeypots that simultaneously emulate multiple industrial protocols. We observe in

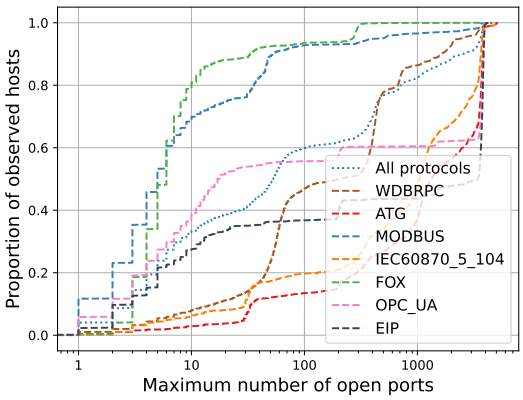

Figure 14. CDFs of the open port count of suspected high- and medium-confidence honeypots running the top 7 ICS protocols (April 2024).

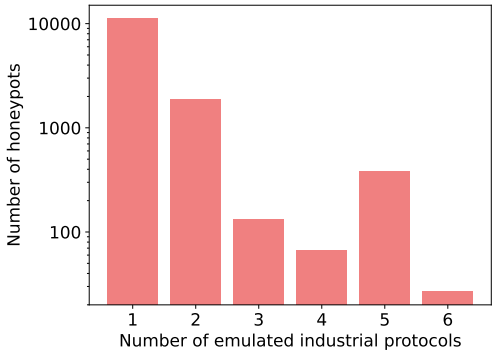

Figure 15. Number of honeypots by emulated industrial protocol count (April 2024).

Figure 15 that the vast majority of suspected honeypots run only one ICS protocol. However, we also found a significant number, 7,668 honeypots, that emulate multiple application-layer industrial protocols. This is observed, for example, in the default setup of popular honeypot software like Conpot [26] and T-Pot [44]. We found honeypots emulating up to 6 industrial protocols on the same IP address.

We focus on pairs of ICS protocols that are commonly hosted together on the same devices. In the case of real devices, we did not find any such pairs. In contrast, in the case of suspected honeypots, we notice some pairs of protocols which seem to typically appear together very often in our study. For example, 79% of IEC 60870-5-104 honeypots also host ATG. This is unexpected, as IEC 60870-5-104 is a protocol used mainly in the power grid, while ATG is used in gas pumps; it is unusual to have both protocols running on the same ICS. Considering all hosts in our study, regardless of classification, we find 1,485 that host both ATG and IEC 60870-5-104. All of them turned out to be classified as honeypots with high confidence; they all match the GasPot fingerprint. Around three-quarters of these hosts are on datacenter networks, spread across a large number of ASes and countries. 1,475 (99%) of them have more than 10 open ports, and 1,408 (95%) have more than 30. We attribute this correlation to honeypot software emulating multiple protocols. While GasPot does not support IEC 60870-5-104, this honeypot software likely uses GasPot as one of its components. Other such pairs may also give hints to multi-protocol honeypots, such as EIP on PRO-CON-OS hosts (95%), DNP3 on PRO-CON-OS hosts (94%), EIP on DNP3 hosts (88%), ATG on DNP3 hosts (85%), and OPC UA on DNP3 hosts (83%). Such correlations between the presence of unrelated protocols on the same host could be used to infer proprietary honeypot families in the wild.

6.7. Oddballs

In the process of validating suspected honeypots, we came across a number of unconventional honeypot families. For example, we discovered 22 hosts, most of which are on the same autonomous system, that appear to be running the same honeypot software. These honeypots have thousands of open ports, out of which Censys is able to identify 34 protocols after performing application-layer handshakes. The identified services include MySQL, Bitcoin, ElasticSearch, TeamViewer, and more than 600 HTTP services, among dozens more, listed in Appendix D.1. This is highly unusual, as those protocols are not related to each other in any way and many of them have no reason to be present on an ICS device. On port 80, we found an HTTP server hosting a very strange page. Its HTML response is 3,950 lines long and contains strings which we hypothesize are used to attract automated vulnerability scanners, as many of them have nothing to do with HTML. They appear to be simulating the expected responses of successful exploitation of various types of vulnerable devices and were similar to the *Anglerfish* honeypot family described in [85]. Appendix D.2 contains some such examples. Grouping hosts by the HTTP response hashes reported by Censys, we found at least 10 more clusters of up to 33 unique IPs each. These clusters serve a similar set of protocols to the honeypot described above and host similar albeit differently structured web pages.

We decided to look into other hosts with a large number of open ports. We investigated the hosts with at least 3,000 different open ports, which uncovered 1,233 unique IPs across 204 different ASes. We found 180 on AS14061 (Digital Ocean), 132 on AS63949 (Akamai), 116 on AS16509 (Amazon), and 87 on AS45102 (Alibaba), among others. Out of the top 20 ASes with the highest number of such hosts, 18 were hosting providers, 1 was an academic network, and 1 was owned by a business.

We conclude that such honeypots may be part of advanced honeypot families with the ability to emulate a wide range of protocols, including ICS. Their behaviour is much different to that of known ICS honeypots like Conpot or the Snap7 framework, which only support a small set of application-layer protocols.

7. Discussion

A high number of honeypots. Our study shows a surprisingly high proportion of honeypots vs. real devices, especially within some countries or protocols. Previous ICS studies did not consider honeypots in their methodologies, or only did rudimentary classification by fingerprinting only specific honeypots [12]–[16], [73]. They may well misidentify honeypots as real and their results may be biased. Our study challenges these findings, as honeypots

seem to be more widely deployed and there are striking differences across regions, networks, and protocols.

Improving stealthiness of honeypots. In addition to signatures, we are able to identify honeypots based on network information or a high number of open ports. Malicious actors could use similar approaches to avoid targeting honeypots and exposing their tactics, techniques, and procedures (TTPs). If identified, honeypots are blacklisted by adversaries and their data collection value is reduced significantly [86]. Our methodology enables honeypot operators to check whether their honeypot deployments can be identified as such and improve their stealthiness and data collection capabilities.

Sophisticated honeypots. We find several highly sophisticated honeypots with the ability to emulate dozens of protocols. We found little public information or previous research on such honeypot families. The limited set of signatures available today can only infer a very small number of the honeypots we discovered in our study. Such signatures are based on known ICS honeypots and are not effective at discovering sophisticated or unknown honeypots.

Differences in measurements. In the studied measurements collected in 2024 and 2025, we notice that the number of real and honeypot hosts within up to a three-month observation window does not increase dramatically. However, it is not yet known if this holds for longer periods. Our tools and methodology can be used to continuously monitor the state of exposed industrial control systems and honeypots that emulate them on the Internet. This is a step towards automatically tracking changes that are important for Internet researchers, policymakers, and engineers.

Limitations. Our methodology relies largely on heuristics for honeypot identification. Heuristics are not perfect — there are likely false positives and false negatives. For example, if a honeypot is on an enterprise network, has only one open port, and does not match any known signature, our methodology would misclassify it as a real device. Furthermore, honeypots could be employed by critical infrastructure operators as part of intrusion detection systems (IDS). However, our methodology would likely be unable to recognize such honeypots based on network information, as they would be placed within critical infrastructure networks and likely have few open ports. Moreover, our methodology cannot find ICS hosts within private networks; it only aims to detect hosts on the public Internet. Most known scanners, like Censys, comply with "opt-out" requests by network providers to not scan their networks [87]. The presented results depend on the reliability of third-party information, i.e., the completeness of Censys data and and the accuracy of IPinfo. We also acknowledge that the operation and configuration of Network Address Translation (NAT) can lead to significant misclassifications or underestimation of the number of exposed ICS devices. For these reasons, the numbers we provide for exposed industrial systems and honeypots are a lower bound.

An alternative to collect raw data would be to utilize publicly available scanning tools like NMap [88] or ZMap [9]. NMap is a "vertical" scanner that scans for open ports of a given host. However, it is intrusive as it may send up to thousands of packets [89] to profile a host. It does not scale well for Internet-wide scanning. On the other hand, ZMap is a "horizontal" scanner — it efficiently and stealthily scans all hosts in a given network range for specific open ports and can scale well for the entire IPv4 range. However, when multiple ports are scanned from a single host, the scanning activity can be detected. Thus, accurate scans with ZMap would require many vantage points to be comparable with those provided by Censys, which operates distributed servers, and reduce bias [64]. Hence, despite this limitation, our analysis offers a large-scale state of deployment of exposed industrial control systems and honeypots in the wild and is reproducible as the raw data is accessible for research purposes.

Best practices for securing exposed devices. ICS protocols often support no authentication at all, which makes securing devices very difficult [9]. In addition, even authenticated services can have configuration errors or software vulnerabilities. This makes any public exposure very dangerous. Even if every known vulnerability is patched, a zero-day vulnerability can occur at any time. Our best-practice recommendation for securing exposed ICS devices is to isolate them from the public Internet. This can be done by air-gapping critical hosts when possible, or via the use of Virtual Private Networks (VPNs) with strong authentication. Moreover, monitoring tools such as Attack Surface Management (ASM) services [90], [91] can enable network operators to automatically detect inadvertently exposed devices and quickly take action.

8. Conclusion

In this study, we considered 17 widely used industrial control protocols and took advantage of application-layer scanning for the entire IPv4 space on all ports. We developed a methodology that uncovered around 150 thousand exposed Internet-facing devices hosting industrial control protocols. In April 2024, we identified 15% of the exposed hosts as suspected honeypots (and 25% in January 2025), with two-thirds classified with high or medium confidence. This high percentage could not be uncovered with previously reported methods that relied primarily on active scanning for open ports and a limited number of ICS fingerprinting signatures. Our analysis shows striking differences between regions, types of networks, and ICS protocols when it comes to the discovery of exposed ICS and related honeypots. It also uncovers a high number of ICS honeypots that are not fingerprintable by known signatures using different heuristics relying on network information and open ports. Thus, our study challenges previous work that reports the demographics of exposed ICS and honeypots.

As part of our future research agenda, we plan to generate new and more reliable ICS and honeypot signatures. We will continue analyzing regular scan snapshots to detect additional exposed ICS devices and honeypots, as their IPs may be dynamic or new ones are deployed. We will continue assessing the hygiene of critical infrastructures that rely on ICS, e.g., manufacturing, power grid, and government, and increase awareness to the various stakeholders. We also plan to uncover exposed industrial control systems and honeypots in the IPv6 address space, which is challenging as IPv6 hit lists are inaccurate or irregularly updated.

Acknowledgments

The authors extend their gratitude to Harm Griffioen and Enrico Bassetti for their feedback on an earlier version of this paper and the anonymous reviewers for their constructive comments and suggestions. They also want to thank Censys for providing research access to their active measurement datasets and IPinfo for providing research access to their geolocation databases. This work was supported by the European Commission under the Horizon Europe Programme as part of the projects SafeHorizon (Grant Agreement #101168562) and RECITALS (Grant Agreement #101168490). Any opinions and conclusions expressed in this article are those of the authors and do not necessarily reflect the official opinion of the European Union.

References

[1] Siemens. (2019) S7 Communication between SIMATIC S7-1200 and SIMATIC S7-300. [Online]. Available: https://cache.industry.siemens.com/dl/files/951/92269951/att_54641/v5/s7communication_s7-1200_s7-300_en.pdf

[2] modbustools.com, "Modbus Protocol Description," 2024, https://www.modbustools.com/.

[3] bacnet.org, "About the BACnet standard," 2024, https://bacnet.org/about-bacnet-standard/.

[4] sk4ld. (2006) INTERNATIONAL STANDARD IEC 60870-5-104. [Online]. Available: https://webstore.iec.ch/preview/info_iec60870-5-104%7Bed2.0%7Den_d.pdf

[5] D. Kushner, "The real story of stuxnet," *IEEE Spectrum*, vol. 50, no. 3, pp. 48–53, 2013.

[6] D. Ltd. (2017) CRASHOVERRIDE: Analysis of the threat to electric grid operations. [Online]. Available: www.dragos.com/blog/crashoverride/CrashOverride-01.pdf

[7] K. E. Hemsley and R. E. Fisher, "History of Industrial Control System Cyber Incidents," 2018, Idaho National Laboratory – Prepared for the U.S. Department of Energy, Office of Nuclear Energy, Under DOE Idaho Operations Office, Contract DE-AC07-05ID14517 .

[8] E. Kovacs. (2023) 670 ICS Vulnerabilities Disclosed by CISA in First Half of 2023: Analysis. https://www.securityweek.com/670-ics-vulnerabilities-disclosed-by-cisa-in-first-half-of-2023-analysis/.

[9] Z. Durumeric, E. Wustrow, and J. A. Halderman, "ZMap: Fast Internet-Wide Scanning and its Security Applications," in *USENIX Security Symposium*, 2013.

[10] M. Dahlmanns, J. Lohmöller, J. Pennekamp, J. Bodenhausen, K. Wehrle, and M. Henze, "Missed opportunities: measuring the untapped TLS support in the industrial internet of things," in *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security*, 2022, pp. 252–266.

[11] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman, "A Search Engine Backed by Internet-Wide Scanning," in *22nd ACM Conference on Computer and Communications Security*, 2015.

[12] J. M. Ceron, J. J. Chromik, J. Santanna, and A. Pras, "Online Discoverability and Vulnerabilities of ICS/SCADA Devices in the Netherlands," arXiv, https://arxiv.org/abs/2011.02019, Nov 2020.

[13] M.-R. Zamiri-Gourabi, A. R. Qalaei, and B. A. Azad, "Gas what?: I can see your GasPots. Studying the fingerprintability of ICS honeypots in the wild," in *ICSS: Proceedings of the Fifth Annual Industrial Control System Security (ICSS) Workshop*. San Juan, PR: Association for Computing Machinery, Dec 2019, pp. 30–37. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3372318.3372322

[14] K. Wilhoit and S. Hilt. (2015) The Little Pump Gauge That Could: Attacks Against Gas Pump Monitoring Systems. https://www.blackhat.com/docs/us-15/materials/us-15-Wilhoit-The-Little-Pump-Gauge-That-Could-Attacks-Against-Gas-Pump-Monitoring-Systems.pdf.

[15] Conpot, "template.xml," 2015, https://github.com/mushorg/conpot/blob/f0e6925fb9632172922abe41b293d7ee438fa60b/conpot/templates/default/template.xml.

[16] Snap7, "server.c," 2015, https://github.com/SCADACS/snap7/blob/f6ff90317ca5d54250f4dcd29209689a74e26d82/examples/plain-c/server.c.

[17] Censys. (2024) Exposure Management and Threat Hunting Solutions — Censys. [Online]. Available: https://censys.com/

[18] L. Izhikevich, R. Teixeira, and Z. Durumeric, "Predicting IPv4 Services Across All Ports," in *ACM SIGCOMM*, 2022.

[19] Censys. (2021) New Universal Internet DataSet Improves Breadth, Depth and Frequency of Scanning. [Online]. Available: https://censys.com/new-universal-dataset-improves-visibility/

[20] L. Izhikevich, R. Teixeira, and Z. Durumeric, "Lzr: Identifying unexpected internet services," in *30th USENIX Security Symposium (USENIX Security)*, 2021.

[21] Z. Durumeric, D. Adrian, P. Stephens, E. Wustrow, and J. A. Halderman, "Ten Years of ZMap," in *ACM Internet Measurement Conference*, 2024.

[22] R. Labs. (2021) Project Sonar. [Online]. Available: https://opendata.rapid7.com/

[23] Censys. (2023) Raising the Bar on Internet Coverage: Predictive Scanning Takes The Censys Internet Map to the Next Level. https://censys.com/raising-the-bar-on-internet-coverage-predictive-scanning-takes-the-censys-internet-map-to-the-next-level/.

[24] ——. (2025) Scanning FAQ. [Online]. Available: https://docs.censys.com/docs/scanning-faq/

[25] H. Griffioen, G. Koursiounis, G. Smaragdakis, and C. Doerr, "Have you SYN me? Characterizing Ten Years of Internet Scanning," in *ACM Internet Measurement Conference*, 2024.

[26] Honeynet Project, "Conpot," 2023, http://conpot.org/.

[27] "Profibus," 2024, https://www.profibus.com/.

[28] rtautomation.com, "A little background on Ethernet/IP," 2024, https://www.rtautomation.com/technologies/ethernetip/.

[29] tridium.com, "Niagara Open Automation Solutions," 2024, https://www.tridium.com/.

[30] Wind River, "Wind River Linux: The Market-Leading Commercial Embedded Linux," 2024, https://www.windriver.com/.

[31] North Dakota UST Operator Training Program. (2012) Automatic Tank Gauging (ATG). [Online]. Available: https://secure.apps.nd.gov/doh/operator/Training/OperatorTraining_ATG.pdf

[32] CODESYS GmbH. (2024) CODESYS Group. [Online]. Available: https://www.codesys.com/

[33] Neuron Team. (2023) Omron FINS Protocol: The Basics & Its Benefits of Bridging to MQTT. [Online]. Available: https://www.emqx.com/en/blog/omron-fins-protocol

[34] OPC Foundation. (2024) Unified Architecture. [Online]. Available: https://opcfoundation.org/about/opc-technologies/opc-ua/

[35] Dnp.org. (2024) Overview Of DNP3 Protocol. [Online]. Available: https://www.dnp.org/About/Overview-of-DNP3-Protocol

[36] Phoenix Contact. (2019) PC WORX BASIC-PRO LIC - Programming software. [Online]. Available: https://www.phoenixcontact.com/en-pc/products/software-pc-worx-basic-pro-lic-2985259

[37] Plant Automation. (2024) ProConOS. [Online]. Available: https://www.plantautomation.com/doc/proconos-0001

[38] Typhoon HIL Documentation. (2023) IEC 61850 MMS Protocol. [Online]. Available: https://www.typhoon-hil.com/documentation/typhoon-hil-software-manual/References/iec_61850_mms_protocol.html

[39] IPESOFT. (2020) General Electric SRTP protocol. [Online]. Available: https://doc.ipesoft.com/display/D2DOCV12EN/General+Electric+SRTP+protocol

[40] FieldComm Group. (2024) HART-IP Explained. [Online]. Available: https://www.fieldcommgroup.org/technologies/HART-IP/explained

[41] A. V. Serbanescu, S. Obermeier, and D.-Y. Yu, "ICS Threat Analysis Using a Large-Scale Honeynets," in *3rd International Symposium for ICS & SCADA Cyber Security Research 2015 (ICS-CSR 2015) (ICS-CSR)*. Ingolstadt, Germany: BCS Learning & Development Ltd., Sep 2015. [Online]. Available: https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/ICS2015.3

[42] GreyNoise. (2025) GreyNoise Intelligence — Real-Time Intelligence For Modern Threats. [Online]. Available: https://www.greynoise.io/

[43] C. Munteanu, S. J. Saidi, O. Gasser, G. Smaragdakis, and A. Feldmann, "Fifteen Months in the Life of a Honeyfarm," in *Proceedings of ACM Internet Measurement Conference (IMC) 2023*, Montreal, QC, Canada, October 2023.

[44] Deutsche Telekom Security GmbH and M. Ochse, "T-Pot," Apr 2022. [Online]. Available: https://github.com/telekom-security/tpotce

[45] MushMush Foundation, "Glutton," 2023, https://github.com/mushorg/glutton.

[46] S. Lau, J. Klick, S. Arndt, and V. Roth, "Towards Highly Interactive Honeypots for Industrial Control Systems," 2016, https://www.scadacs.org/pubs/2016_plc_honeypot_poster.pdf.

[47] D. I. Buza, F. Juhász, G. Miru, M. Félegyházi, and T. Holczer, "CryPLH: Protecting Smart Energy Systems from Targeted Attacks with a PLC Honeypot," in *International Workshop on Smart Grid Security, SmartGridSec 2014*. Munich, Germany: Springer, Jan 2014. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-10329-7_12

[48] F. Xiao, E. Chen, and Q. Xu, "S7commTrace: A High Interactive Honeypot for Industrial Control System Based on S7 Protocol," in *International Conference on Information and Communications Security (ICICS 2017)*. Springer, Apr 2017. [Online]. Available: https://www.springerprofessional.de/en/s7commtrace-a-high-interactive-honeypot-for-industrial-control-s/15646288

[49] E. López-Morales, C. Rubio-Medrano, A. Doupé, Y. Shoshitaishvili, R. Wang, T. Bao, and G.-J. Ahn, "HoneyPLC: A Next-Generation Honeypot for Industrial Control Systems," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS '20)*, October 2020, pp. 279–291, published: 02 November 2020.

[50] C. Ding, J. Zhai, and Y. Dai, "An Improved ICS Honeypot Based on SNAP7 and IMUNES," in *International Conference on Cloud Computing and Security (ICCCS 2018)*, 2018, pp. 303–313, first Online: 01 November 2018.

[51] ArtWachowski, "dnp3pot," 2020, https://github.com/ArtWachowski/dnp3pot.

[52] K. Kołtyś and R. Gajewski, "SHaPe: A Honeypot for Electric Power Substation," *Journal of Telecommunications and Information Technology*, no. 4, 2015.

[53] Niels Provos. (2023) Developments Of The Honeyd Virtual Honeypot. [Online]. Available: https://honeyd.org

[54] sk4ld. (2023) GridPot. [Online]. Available: https://github.com/sk4ld/gridpot

[55] S. Srinivasa, J. M. Pedersen, and E. Vasilomanolakis, "Gotta Catch 'em All: A Multistage Framework for Honeypot Fingerprinting," *Digital Threats*, vol. 4, no. 3, oct 2023. [Online]. Available: https://doi.org/10.1145/3584976

[56] A. Mirian, Z. Ma, D. Adrian, M. Tischer, T. Chuenchujit, T. Yardley, R. Berthier, J. Mason, Z. Durumeric, J. A. Halderman, and M. Bailey, "An Internet-wide view of ICS devices," in *2016 14th Annual Conference on Privacy, Security and Trust (PST)*. Auckland, New Zealand: IEEE, Dec 2016. [Online]. Available: https://ieeexplore.ieee.org/xpl/conhome/7899233/proceeding

[57] Bitsight Security Research, "Bitsight identifies nearly 100,000 exposed industrial control systems," 2023, https://www.bitsight.com/blog/bitsight-identifies-nearly-100000-exposed-industrial-control-systems.

[58] T. Basin, Y. Sade, and Y. Harel, "Research: Nearly 70,000 Sensitive Industrial Control Systems Exposed," 2021, https://www.otorio.com/blog/ics-exposures-research-blog/.

[59] Shodan. (2024) Shodan Search Engine. [Online]. Available: https://www.shodan.io/about/products

[60] R. Yaben, N. Lundsgaard, J. August, and E. Vasilomanolakis, "Towards identifying neglected, obsolete, and abandoned IoT and OT devices," in *8th Network Traffic Measurement and Analysis Conference*. IFIP, 2024.

[61] B. Wang, X. Li, L. P. de Aguiar, D. S. Menasche, and Z. Shafiq, "Characterizing and modeling patching practices of industrial control systems," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 1, no. 1, Jun 2017.

[62] Shodan, "Honeyscore." [Online]. Available: https://honeyscore.shodan.io/

[63] Y. Wu, S. Song, J. Zhuge, T. Yin, T. Li, J. Zhu, G. Guo, Y. Liu, and J. Hu, "ICScope: Detecting and Measuring Vulnerable ICS Devices Exposed on the Internet," in *Information Systems Security and Privacy. ICISSP 2021, ICISSP 2022. Communications in Computer and Information Science*. Ingolstadt, Germany: Springer, Cham, Jul 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-37807-2_1

[64] G. Wan, L. Izhikevich, D. Adrian, K. Yoshioka, R. Holz, C. Rossow, and Z. Durumeric, "On the Origin of Scanning: The Impact of Location on Internet-wide Scans," in *Proceedings of the ACM Internet Measurement Conference*, 2020, pp. 662–679.

[65] P. Gigis, M. Calder, L. Manassakis, G. Nomikos, V. Kotronis, X. Dimitropoulos, E. Katz-Bassett, and G. Smaragdakis, "Seven Years in the Life of Hypergiants' Off-Nets," in *Proceedings of ACM SIGCOMM 2021*, Virtual Event, August 2021.

[66] IPinfo. (2024) The trusted source for IP address data. [Online]. Available: https://ipinfo.io/

[67] M. J. Freedman, M. Vutukuru, N. Feamster, and H. Balakrishnan, "Geographic Locality of IP Prefixes," in *ACM IMC*, 2005.

[68] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe, "Towards IP geolocation using delay and topology measurements," in *ACM IMC*, 2006.

[69] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye, "IP Geolocation Databases: Unreliable?" *ACM CCR*, vol. 41, no. 2, 2011.

[70] B. Huffaker, M. Fomenkov, and k. claffy, "Geocompare: A Comparison of Public and Commercial Geolocation Databases," CAIDA, Tech. Rep., May 2011, http://www.caida.org/publications/papers/2011/geocompare-tr/.

[71] J. Zirngibl, L. Steger, P. Sattler, O. Gasser, and G. Carle, "Rusty clusters? dusting an ipv6 research foundation," in *Proceedings of the 22nd ACM Internet Measurement Conference*, 2022, pp. 395–409.

[72] Snap7, "Snap7 Homepage," 2016, https://snap7.sourceforge.net/.

[73] Censys. (2023) Unmasking Deception: Honeypot and Red Herring in Network Security. [Online]. Available: https://censys.com/red-herrings-and-honeypots/

[74] S. Maesschalck, V. Giotsas, and N. Race, "World wide ics honeypots: A study into the deployment of conpot honeypots," in *Industrial Control System Security Workshop*, 2021.

[75] Conpot, "guardian_ast_server.py," 2022, https://github.com/mushorg/conpot/blob/f0e6925fb9632172922abe41b293d7ee438fa60b/conpot/protocols/guardian_ast/guardian_ast_server.py.

[76] H. Griffioen, K. Oosthoek, P. van der Knaap, and C. Doerr, "Scan, Test, Execute: Adversarial Tactics in Amplification DDoS Attacks," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '21. Association for Computing Machinery, 2021, p. 940–954. [Online]. Available: https://doi.org/10.1145/3460120.3484747

[77] A. Jicha, M. Patton, and H. Chen, "SCADA honeypots: An in-depth analysis of Conpot," in *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, 2016, pp. 196–198.

[78] M. Mladenov, "Detection of critical infrastructure devices on the public Internet," June 2023.

[79] A. Kyriakou and N. Sklavos, "Container-Based Honeypot Deployment for the Analysis of Malicious Activity," in *2018 Global Information Infrastructure and Networking Symposium (GIIS)*, 2018.

[80] Censys. (2021) Moving Beyond the Noise by Filtering Internet Pseudo Services. [Online]. Available: https://censys.com/beyond-noise-by-filtering-pseudo-services/

[81] B. Stock, G. Pellegrino, C. Rossow, M. Johns, and M. Backes, "Hey, you have a problem: On the feasibility of Large-Scale web vulnerability notification," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug 2016, pp. 1015–1032.

[82] Censys. (2025) Labels. [Online]. Available: https://docs.censys.com/docs/ls-labels

[83] Nmap. (2020) s7-info.nse. [Online]. Available: https://github.com/nmap/nmap/blob/63e63bd99976ae3700d4dc14e4510260a542b297/scripts/s7-info.nse#L198

[84] P. Richter, F. Wohlfart, N. Vallina-Rodriguez, M. Allman, R. Bush, A. Feldmann, C. Kreibich, N. Weaver, and V. Paxson, "A multi-perspective analysis of carrier-grade nat deployment," in *Proceedings of the 2016 Internet Measurement Conference*, ser. IMC '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 215–229. [Online]. Available: https://doi.org/10.1145/2987443.2987474

[85] n0mad. (2020) Trawling for Fishermen - Investigating a Chinese Honeynet. [Online]. Available: https://www.secjuice.com/trawling-for-fishermen-or-investigating-a-chinese-honeynet/

[86] A. Vetterl and R. Clayton, "Bitter Harvest: Systematically Fingerprinting Low- and Medium-interaction Honeypots at Internet Scale," in *12th USENIX Workshop on Offensive Technologies (WOOT 18)*. Baltimore, MD: USENIX Association, Aug 2018. [Online]. Available: https://www.usenix.org/conference/woot18/presentation/vetterl

[87] Censys, "Opt Out of Data Collection," https://support.censys.io/hc/en-us/articles/360043177092-Opt-Out-of-Data-Collection, 2024.

[88] A. J. Bennieston, "NMAP - A Stealth Port Scanner," 2004.

[89] T. Albakour, O. Gasser, R. Beverly, and G. Smaragdakis, "Illuminating Router Vendor Diversity Within Providers and Along Network Paths," in *Proceedings of ACM Internet Measurement Conference*, October 2023.

[90] Palo Alto Networks. (2024) What Is Attack Surface Management? [Online]. Available: https://www.paloaltonetworks.com/cyberpedia/what-is-attack-surface-management

[91] Censys. (2024) Attack Surface Management — Censys. [Online]. Available: https://censys.com/attack-surface-management/

Appendix

Appendix A. Data Availability

We have shared the code used in this research in a public GitHub repository for reproducibility purposes: https://github.com/martinmladenov/ICS-Honeypots.

On request, we can share aggregated data (e.g., per AS or per country) used for plot generation. We are not permitted to share raw data, as this data is intellectual property of Censys and IPinfo. Such data can be requested directly from Censys and IPinfo for academic use.

We can privately share the data we collected for validation upon request. We have decided not to make it publicly available due to its potential for abuse by malicious parties: if published, IPs of discovered ICS devices could be used as targets by malicious actors. Furthermore, revealing the IP addresses of discovered honeypots could reduce their effectiveness in detecting and analyzing malicious activity.

Appendix B. Aggregated Data

Aggregated information about the detected hosts in August 2024, October 2024, and January 2025 is shown in Table 7. Information about the hosts in each classification category in January 2025 can be seen in Table 8.

Appendix C. Protocol Distribution per Country

Figures 16 and 17, respectively, show the distribution of exposed industrial protocols per country in the case of real devices and of honeypots.

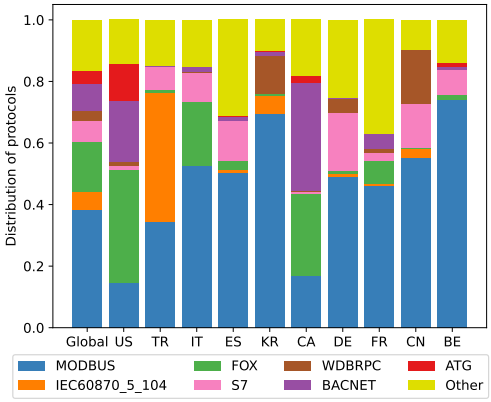

Figure 16. Distribution of exposed industrial protocols in real ICS devices, per country (April 2024).

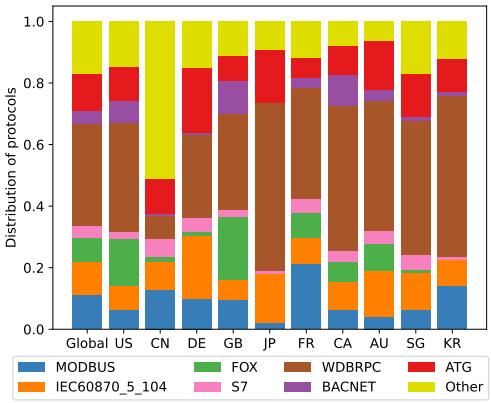

Figure 17. Distribution of industrial protocols used by suspected honeypots, per country (April 2024).

Appendix D. Anglerfish-like Honeypot

We found numerous sophisticated honeypots with thousands of open ports, emulating a variety of protocols. They appear to be very similar to the Anglerfish honeypot family. This appendix includes information about one such honeypot — namely, the identified protocols (Appendix D.1) and some unusual snippets from the HTML page on port 80 (Appendix D.2).

TABLE 7. PROTOCOLS AND AGGREGATED DATA FROM AUGUST 2024, OCTOBER 2024, AND JANUARY 2025.

	August 6th, 2024						October 23rd, 2024						January 28th, 2025					
	Hosts		ASes		Countries		Hosts		ASes		Countries		Hosts		ASes		Countries	
Protocol	Real	HP	Real	HP	Real	HP	Real	HP	Real	HP	Real	HP	Real	HP	Real	HP	Real	HP
MODBUS	43,191	4,244	1,931	847	127	70	40,062	4,467	1,786	841	131	76	42,661	4,638	1,888	756	124	76
FOX	19,534	2,667	1,062	610	72	54	19,202	2,656	1,005	616	70	57	19,240	2,559	1,012	596	66	54
BACNET	10,259	1,343	923	448	85	63	12,213	1,877	990	624	94	88	13,112	2,000	1,033	609	87	82
EIP	9,730	1,166	578	223	81	64	9,067	1,094	542	200	80	66	9,790	1,413	521	185	74	66
IEC60870_5_104	7,335	2,192	227	446	55	92	6,704	1,847	195	410	56	90	7,155	1,608	217	428	53	86
S7	7,546	859	601	245	78	59	7,414	890	578	257	78	59	7,641	632	586	209	73	53
ATG	5,463	2,426	474	427	35	93	5,176	2,120	447	391	36	92	7,769	2,296	543	418	48	92
WDBRPC	6,572	7,863	730	812	107	87	7,945	15,558	721	864	109	89	8,669	26,228	631	747	103	84
CODESYS	2,674	253	292	127	54	28	2,511	256	285	135	52	31	2,579	256	293	122	50	31
FINS	2,503	295	250	96	58	33	2,238	278	243	105	59	32	4,015	1,148	270	109	59	53
OPC_UA	1,511	1,362	371	281	73	65	1,585	1,396	336	298	68	71	1,810	1,814	347	284	69	66
DNP3	556	676	83	52	28	41	578	685	79	46	30	41	599	1,054	90	57	29	43
PCWORX	455	19	59	10	16	8	450	17	49	8	18	8	472	22	52	9	15	9
PRO_CON_OS	39	604	14	24	7	38	35	610	11	18	6	40	33	943	10	28	5	43
MMS	51	28	18	13	13	8	41	17	11	9	9	7	52	25	14	11	12	8
GE_SRTP	48	7	28	6	15	5	46	11	25	9	14	6	42	10	21	8	10	7
HART	6	6	3	1	1	1	5	6	3	1	1	1	7	5	4	1	1	1
Total	109,871	20,812	3,897	2,489	164	119	107,943	28,840	3,702	2,591	166	126	116,764	39,866	3,833	2,445	158	124

TABLE 8. NUMBER OF HOSTS IN EACH CLASSIFICATION CATEGORY AND THE NUMBER OF ASES AND COUNTRIES WITH AT LEAST ONE HOST IN A GIVEN CATEGORY (JANUARY 28TH, 2025).

	Hosts				Autonomous Systems				Countries			
	Real	Honeypots			Real	Honeypots			Real	Honeypots		
Protocol		*Low*	*Medium*	*High*		*Low*	*Medium*	*High*		*Low*	*Medium*	*High*
MODBUS	42,661	2,537	2,097	4	1,888	459	356	3	124	63	62	4
WDBRPC	8,669	992	25,236	0	631	214	581	0	103	56	74	0
FOX	19,240	1,660	899	0	1,012	382	245	0	66	47	41	0
BACNET	13,112	1,325	675	0	1,033	381	257	0	87	73	54	0
EIP	9,790	295	1,068	50	521	95	76	25	74	33	48	26
ATG	7,769	81	962	1,253	543	32	44	364	48	4	45	83
IEC60870_5_104	7,155	114	435	1,059	217	39	87	338	53	19	42	81
S7	7,641	182	374	76	586	98	84	34	73	31	40	29
FINS	4,015	130	1,018	0	270	56	58	0	59	24	45	0
OPC_UA	1,810	351	1,463	0	347	151	142	0	69	49	58	0
CODESYS	2,579	175	81	0	293	83	46	0	50	29	17	0
DNP3	599	31	1,023	0	90	15	42	0	29	11	42	0
PRO_CON_OS	33	1	942	0	10	1	27	0	5	1	43	0
PCWORX	472	12	10	0	52	6	4	0	15	6	4	0
MMS	52	3	22	0	14	3	8	0	12	2	6	0
GE_SRTP	42	10	0	0	21	8	0	0	10	7	0	0
HART	7	5	0	0	4	1	0	0	1	1	0	0
Total	116,764	7,351	31,199	1,316	3,833	1,204	1,261	385	158	98	94	86

D.1. Encountered Protocols

- AMQP
- BITCOIN (on 2 ports)
- DNP3
- ELASTICSEARCH
- HTTP (on 641 ports)
- MDNS
- MONERO P2P
- MQTT (on 2 ports)
- MYSQL
- NETIS
- PCOM
- PPTP
- REDIS
- SIP (on 2 ports)
- SOCKS
- TEAM VIEWER
- UBIQUITI
- Unidentified (3840)
- ATG
- DICOM (on 5 ports)
- EIP
- FTP
- IPP (on 13 ports)
- MEMCACHED
- MONGODB
- MSSQL
- NETBIOS
- OPC UA
- POP3
- PRO CON OS
- RTSP
- SMTP (on 2 ports)
- SSH (on 3 ports)
- TELNET (on 2 ports)
- VNC

D.2. HTML Page on Port 80

The HTML source of the index page of the HTTP server on port 80 contained some unusual elements. The Snippets 1-9 below show some interesting examples.

Snippet 1. Numerous `<meta>` elements. Some appear to be taken from a GitLab instance, some from a VoIP product.

```
<!DOCTYPE html>
<html>
<head>
<meta http-equiv="Content-Type"
content="text/html; charset=UTF-8" />
<meta http-equiv="X-UA-Compatible"
content="IE=edge">
<meta http-equiv="Pragma" content="no-cache" />
<meta charset="utf-8">
<meta content="IE=edge"
http-equiv="X-UA-Compatible">
<meta content="object" property="og:type">
<meta content="GitLab" property="og:site_name">
<meta content="Help" property="og:title">
```

```
<meta content="GitLab Community Edition"
property="og:description">
<meta content="summary" property="twitter:card">
<meta content="Help" property="twitter:title">
<meta content="GitLab Community Edition"
property="twitter:description">
<meta content="GitLab Community Edition"
name="description">
<meta content="#474D57" name="theme-color">
<meta content="#30353E"
name="msapplication-TileColor">
<meta name="csrf-param"
content="authenticity_token" />
<meta name="csrf-token"
content="8dcb74a64dc984fb9abe3e7c201f810d9ec
90ed8e4c970c632b49be7fed5240a23==" />
<meta http-equiv="Content-Type"
content="text/html; charset=utf-8"/>
<meta http-equiv="expires" content="-1"/>
<meta name="keywords" content="VOS3000, VoIP,
VoIP运营支撑系统, 软交换"/>
<meta name="description" content="VOS3000, VoIP,
VoIP运营支撑系统, 软交换"/>
<meta name="author" content="www.linknat.com, 昆
石网络"/>
<meta name="copyright" content="www.linknat.com,
昆石网络"/>
<link rel="shortcut icon"
href="images/vos3000.ico"/>
```

Snippet 2. The characters `:`, `#`, `>`, and `$` are commonly used in shells.

```
<br>Cisco, Cisco Systems, and the Cisco Systems
logo are registered
trademarks or trademarks of Cisco Systems, Inc.
and/or it's affiliates
in the United States and certain other
countries.
</td>
:
#
>
$
SSH key is good
is not a valid ref and may not be archived
pcPassword2
'&sessionKey=790148060;'
name="sessionKey" value="790148060"
Set-Cookie: loginName=admin
var fgt_lang = /dev/cmdb/sslvpn_websession
php 8.1.0-dev exit
springframework
Tomcat
DEVICE.ACCOUNT=admin
AUTHORIZED_GROUP=1
<uid></uid>
<name>Admin</name>
<usrid></usrid>
<password>admin</password>
<group></group>
cpto /tmp/"root"
```

Snippet 3. Contents of a `/proc/.../smaps` file of a Linux system.

```
X-Content-Powered-By: K2 v2.8.0 (by JoomlaWorks)
007b2000-007c1000 rw-p 00000000 00:00 0
Size: 60 kB
Rss: 52 kB
Pss: 52 kB
Shared_Clean: 0 kB
Private_Clean: 0 kB
Private_Dirty: 52 kB
Referenced: 52 kB
Anonymous: 52 kB
AnonHugePages: 0 kB
Swap: 8 kB
KernelPageSize: 4 kB
MMUPageSize: 4 kB
```

Snippet 4. A reference to a WordPress vulnerability (CVE-2022-1609), and output of the `id` program on a Linux machine.

```
9061-220-EVC
CVE-2022-1609
Hardware:"586"
<pre>
/root
uid=13883(root) gid=13883(root)
groups=13883(root)
uid=13883(rootxx) gid=13883(rootxx)
groups=13883(rootxx)
62318aca2ef2e809a13623715a8aaff4
62318aca2ef2e809
a13623715a8aaff4
muie1976
if('1' == '0' || 'admin' == 'admin')
</pre>
```

Snippet 5. Contents of a `/etc/passwd` file of a Linux system.

```
~~~
root:x:0:0:root:/root:/bin/bash
daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin
bin:x:2:2:bin:/bin:/usr/sbin/nologin
sys:x:3:3:sys:/dev:/usr/sbin/nologin
sync:x:4:65534:sync:/bin:/bin/sync
uucp:x:10:10:uucp:/var/spool/uucp:/usr/sbin/nologin
proxy:x:13:13:proxy:/bin:/usr/sbin/nologin
www-data:x:33:33:www-data:/var/www:/usr/sbin/nologin
backup:x:34:34:backup:/var/backups:/usr/sbin/nologin
nobody:x:65534:65534:nobody:/nonexistent:/usr/sbin/nologin
~~~
```

Snippet 6. A list of `<div>` elements containing ASUS router models.

```
<div class="prod_madelName">RT-AC1900P</div>
<div class="prod_madelName">RT-AC1900U</div>
<div class="prod_madelName">RT-AC3100</div>
<div class="prod_madelName">RT-AC3200</div>
<div class="prod_madelName">RT-AC5300</div>
<div class="prod_madelName">RT-AC56U</div>
<div class="prod_madelName">RT-AC66U B1</div>
<div class="prod_madelName">RT-AC66U_B1</div>
<div class="prod_madelName">RT-AC67U</div>
<div class="prod_madelName">RT-AC68P</div>
<div class="prod_madelName">RT-AC68R</div>
<div class="prod_madelName">RT-AC68U</div>
<div class="prod_madelName">RT-AC68W</div>
<div class="prod_madelName">RT-AC87R</div>
<div class="prod_madelName">RT-AC87U</div>
<div class="prod_madelName">RT-AC88U</div>
<div class="prod_madelName">RT-AX3000</div>
<div class="prod_madelName">RT-AX55</div>
<div class="prod_madelName">RT-AX56U</div>
<div class="prod_madelName">RT-AX58U</div>
<div class="prod_madelName">RT-AX82U</div>
<div class="prod_madelName">RT-AX86U</div>
```

Snippet 7. Various mobile operating systems as well as the contents of the root directory of a Linux machine.

```
1.<a href="PcamEn.htm"><strong>Windows
Mobile/Pocket PC</a></p>
2.<a href="3rd.htm"><strong>Symbian</a></p>
3.<a
href="BlackBerry.htm"><strong>BlackBerry</a></p>
plugins/images/vos.png
var
usr
tmp
sys
srv
service
sbin
run
root
proc
opt
mnt
media
lost+found
lib64
lib
home
etc
dev
da1
boot
bin
Copyright (c) 2015-2020 by Cisco Systems, Inc.
All rights reserved.
```

Snippet 8. Various `<title>` elements, possibly from the web interfaces of home routers.

```
<title>CPPLUS DVR {Web View</title>
<title>OoklaServer</title>
<title>Apache Tomcat/9.0.55</title>
<title>Broadcom ADSL Router,Broadcom single-chip
ADSL router</title>
<title>HK-VT7116 192.168.1.64,Digital Video
Recorder</title>
<title>TK-ONU1P-DUAL</title>
<title>PON Home Gateway</title>
<title>维盟（WayOS）智能路由管理系
统www.wayos.com</title>
<title>NVR308-64E</title>
<title>kiosk_edge_gui</title>
<title>Vantex Panel</title>
<title>cnPilot R190W Login</title>
<title>Cisco RV340 Configuration Utility</title>
<title>Eltex - NTE-RG-1421G-Wac</title>
<title>CAM6082QIR,MAC:00-0F-0D-2A-64-37</title>
<title>Account Suspended</title>
<title>ConfuserEx Online</title>
<title>22NoluKios (build 3249M) - Bilgi</title>
<title>Worktop</title>
<title>Turkcell Güvenli Internet</title>
<title>FTTX Router XPON-1G</title>
<title>Router -> Login</title>
<title>IIS7</title>
<title>XCN News</title>
<title>TransPort WR11 (SN: 754181) Configuration
and Management</title>
<title>Netgear Prosafe Plus Switch</title>
```

Snippet 9. Multiple unrelated `<title>` elements and HTTP `Server` headers.

```
<title>netns-pppoe-converge netdata
dashboard</title>
<title>Arcadia</title>
<title>XWebPlay</title>
<title>Unauthorized</title>
<title>WO-67</title>
<title>Dell SonicWALL - Authentication</title>
<title>Password required</title>
<title>Welcome to CentOS</title>
<title>NETIS RX30,NETIS RX30</title>
<title>4G MIFI</title>

Server: Linux,WEBACCESS/1.0,DIR-860LVer1.07
Server: Werkzeug/2.0.1 Python/3.6.9
Server: QWS
Server: t1-httpd/1.4.43
Server: Indy/10.0.52
Server: Jetty(6.1.26)
Server: LANCOM
Server: waitress
Server: none
Server: Saia PCD3.M5340/1.16.69
Server: WhoAmI
Server: PAM360
Server: TwistedWeb/17.9.0
Server: openresty/1.15.8.3
Server: Apache/1.3.28 (Unix) mod_ssl/2.8.15
OpenSSL/0.9.8d
Server: Resin/4.0.40
Server: webcache
Server: GeoWebServer 5.0.0.0
Server: Hikvision-Webs
Server: huohuo
Server: ioLogik Web Server/1.0
Server: HttpServer
Server: Niagara Web Server/3.8.111
Server: HFS 2.3c
Server: rcell
Server: H3C-CVM
```