ATTACK SURFACE DESCRIPTION LANGUAGE

Dheepak Nalluri

Master of Science

APRIL 22, 2021

DEPARTMENT OF COMPUTER ENGINEERING
Iowa State University

Abstract

Documenting and describing attack surfaces is a common tactic in parts of the industry for various reasons. Some for open design, and others for penetration testing. Either way, there is no standardized methods for documenting attack surfaces. This paper presents an Attack Surface Description Language (ASDL), a way to describe and present a device's attack surfaces. These surfaces can be documented as known or unknown in a way that allows Blackbox testers to use ASDL as well. Complex structures can also be represented with dependencies on lower-level structures. Ease of use and flexibility was also taken into account in the design of ASDL to make it more efficient and less tedious to use.

1) Introduction

1.1) Motivation

Connectivity and customizability are expected by consumers in many different types of electronic devices. As the number and variety of communicating devices grow, so do the diversity in their communication methods. However, the consumer does not always know how a device is communicating or all the ways in which the device can interact with other devices. When left to the manufacturer to define all modes of communication on the device, manufacturers can sometimes leave out vital details/features and vulnerabilities. The office printer, XEROX Docucenter DC 230ST, is an example that had many ports open for unknown reasons and was vulnerable to outside attacks in addition to not having any documentation provided by the manual [1]. Even if some features of a communication device are not secure, providing the consumer with the details of the communications can allow a consumer to understand it and configure it in a way to protect it.

1.2) Solution Requirements

What the solution requires is a standard method of defining each type of communication. The method must encompass all types of electronic communication from wired to wireless and from low level protocols to the application layer. Multiple modes of communication on a device should not hurt the readability/usability of the method. Such a method must be able to adapt to future technologies as well. As for ease of use, the documentation of a product must be easy to navigate and understand by a mildly technical consumer base. It must also be somewhat compact to fit within a reasonably sized manual/package.

1.3) Benefits

There exist many benefits to standardizing such a method for both consumers and companies. One benefit of a standard method is for consumer awareness. Consumers would be able to make educated decisions on what kind of vulnerabilities/attack surfaces could exist in the system. They could also decide on different products by how they are going to use it and what

methods of communication are available. Another benefit is owner configuration and security management. By knowing what communications are being done by a device, an owner could protect the attack surfaces of the device or disable vulnerable surfaces. Pentesting tools would use ASDL and become more autonomous as all attack surfaces and communication methods on a device in a standardized way. Standardization of such a language would force product designers to document and publicly disclose attack surfaces of their system, which will cause those designers to use more open designs as their methods are known.

1.4) Background Work

Very few background works exist that are about standardizing a view of attack surfaces. There are a few related patents and works that are somewhat related to or factor into this paper. One such patent is the 'System and method for electronic communications' held by the Royal Bank of Canada [2]. This patent uses machine learning to learn a systems interfaces and output a graph-based view of its interfaces. However, the graph system does not adapt well to unique interfaces and systems that might exist. It only seems to work well in specific systems. This patent affects this paper because it shows an example of how a graph-based view of a system would work. Another patent is the 'System and method for sorting electronic communications' held by AT&T [3]. This is a patent oriented toward consumer awareness, such that they may refuse certain communications (i.e., ads) and allow others. Allowing consumers to understand products and make an educated decision of their purchase is the goal of this paper, and this patent is another way in which the information could be represented to a consumer. A related work is 'Measuring a System's Attack Surface' by Manadhata and Wing from Carnegie Mellon University [4]. The paper defines formal definitions of attack surfaces and explains how attack surfaces can be found. Since the solution must be somewhat adaptable, the information derived from finding an attack surface using the method from this paper is likely the information that will be used toward the representation of the attack surfaces. The last paper is 'Penetration Analysis of a XEROX Docucenter DC 230ST: Assessing the Security of a Multi-purpose Office Machine' by Daniels, Kuperman, and Spafford [1]. The paper discusses the attack surfaces of an office printer that was not well documented. There existed a lot of TCP and UDP ports that were vulnerable in addition to many physical ports that were not documented. The unsecure ports on the machine caused it to be a very vulnerable piece of hardware that was poorly documented. This is a perfect case of why a standardized language is needed to represent attack surfaces.

1.5) Purpose

The idea that we are proposing in this paper is a specific language that can be standardized to describe all methods of electronic communication that exist on any device. This language will be represented in an XML format and should be concise yet comprehensive enough to cover everything.

Contributions

- A description language that can represent the attack surfaces of any device
- Explain the design choices that satisfy the solution requirements
- Examples of how the description language can be used and represented

Definitions

- Attack Surface – A physical or digital method in which a device can be interacted with or used
- XML – A language that is both machine readable and human readable
- ASDL – Attack Surface Description Language. The XML format language presented in this paper used to describe attack surfaces

2) Description and Methodology

The Attack Surface Description Language, hereby referred to as ASDL, must encompass many different types of communications differing in quantity and complexity. In the following sections, the design decisions are explained in detail.

2.1) Presentation

The overall presentation of ASDL is important as it gives the reader the first impression of whether they would want to read it or not. If ASDL is a cluttered mess, then the reader would have a hard time finding the information that they need. The opposite is then also true as ADSL would not convey enough information. As such, it needs to be well organized and presentable information. Devices can become complex and hold many different methods for interacting with the outside world, leading to an extensive amount of documentation. For this reason, following how databases display their data to users is a good place to start. Two main methods database technologies use are XML and graphs. Graphs are a difficult choice as they get rapidly complex with more items added. The size of a graphical representation would make it hard to fit in any manual or documentation that comes with a device. The graph would be too small to read, broken up over multiple pages, or (most likely) omitted from documentation. XML solves all these problems as it can be rescaled easily to become more readable, and XML can be broken up over multiple pages. In addition, formatting ASDL in XML would make it more readable by automated programs for any purpose. Therefore, ASDL is represented in an XML format.

2.2) Organization

With the format of ASDL solidified, the next issue becomes categorizing and organizing the order in which attack surfaces are displayed. The order matters because it can speed up how a consumer can search for an interface. Not all consumers will be experts in electronic

communications and protocols and as such, would not want to look through tables of protocols being used. However, most consumers will understand what Human Interface Devices, or HIDs, and sensors are on a device. Placing these categories first will increase ASDL's ease of use. The order of the categories are as follows.

- HIDs/Displays
- Sensors/Actuators
- Physical Ports
- Wireless Interfaces
- Wi-Fi/Network
- Bluetooth
- Capabilities
- Misc.

To account for dependencies of any technologies, each row contains a column for ID and Dependency. The ID increases by 1 for each row beginning at ID 1. These IDs keep incrementing between categories and should be unique for each row. The Dependency column then lists the IDs of all the rows that the row immediately depends on in a comma separated list. Note that some technologies can switch between dependencies such as a TCP/IP stack could switch between a wired, Wi-Fi, or mobile connection. These dependencies that are not certain or interchanging are denoted by an underline.

Some other fields in every row are the Name and a Description. The Name is a simple indicator of what that row is about. A name could be as simple as "B button." The main purpose is to allow a reader to quickly peruse and find the object/technology they are looking for. The Description field is a more verbose name field. It could be used to further describe a technology or mention exceptions/quirks with a particular technology.

The last field is the certainty field, simply labeled 'C'. The documentation of certain devices might be more theories or educated guesses on how a device works. This field would remain empty until all elements in the row are certain (even if a field is labeled as unknown), at which point they would mark the column with the 'X' character. This field allows for the documentation of theories or guesses that can later be tested and changed as needed. In the scenario of a manufacturer documenting their product using ASDL, the certainty column can and should be excluded as all rows are certain.

HIDs/Displays

Human input devices and displays tie into the main functionality of many devices on the market. The average consumer would then be most interested in how the device can be physically interacted with. The table is as follows:

ID	Name	Interaction	Dependency	Description	C

The interaction field is the action that the user takes to interact with the item. For example, a button would be “press” while a display would be “visual.” In the scenario that the consumer does not know how to interact with a certain part of a device, this field would clear it up.

Sensors/Actuators

Sensors and Actuators are not directly controlled by the user but can affect how the device works. The average consumer could then try to find an optimal environment or situation in which the device would function. The table is as follows:

ID	Name	Function	Dependency	Description	C

The Function field lists the action that the sensors/actuators can take. A sensor could have a function of sensing distance to a wall in a room, or simply just “ultrasonic.” An actuator would have the description of the movement that is taking places such as ‘rotate glass.’ The Function field’s purpose is meant to be in the description field but is not needed all the time.

Physical Ports

Physical ports exist on about every device to allow for more functionality or ease of use for the consumer. The average consumer still understands what these are and may or may not decide to use them. The table is as follows:

ID	Name	Standardization	Dependency	Description	C

The only added field in this section is Standardization. The reason for this is because a vast majority of physical ports are self-explanatory or for one specific purpose. An IEEE standardization could be listed or just a name. It is possible for the Standardization field to have the exact same contents as the Name field, as long as looking up the Standardization field would lead a consumer to images or a description.

Wireless Interfaces

Wireless Interfaces allow for many IoT devices to function across homes. This section includes how a device wirelessly interacts with other devices but is limited to the hardware aspect of communication. The average consumer does not always understand this section and does not often need to reference it. The table is as follows:

ID	Name	Frequency Range	Dependency	Description	C

The Frequency Range is the range of the frequencies emitted or received by the device. This section could include the entire electromagnetic spectrum but is generally meant to contain infrared and lower frequency communication. The Protocol is the data link layer. An IEEE standardization or method can be listed, if the user can look up the protocol and get related results.

Wi-Fi/Network/Mobile

Connections to the internet or local networks are an important part of ASDL as these have historically been the most vulnerable. Despite their importance, this section starts to get more technical for the average consumer and is put 5th on the list. The table is as follows:

ID	Name	Port(s)	Protocol	Dependency	Description	C

The Protocol field in this section refers to the protocol being run to communicate between two communicating devices. Custom protocols can be listed as ‘Custom’ and described in further detail in the Description field. The Port(s) field lists the port that a service is running on. Multiple ports can be listed in this field such that the table does not become cumbersome or tedious.

Bluetooth

This section can normally be included in the wireless interfaces section, but due to the increasing popularity of wireless Bluetooth devices, they would be best in its own section. However, if Bluetooth becomes less used in the next few years, this section could be removed entirely and replaced with newer technologies. The table is as follows:

ID	Name	Protocol	Dependency	Description	C

There can be many layers of protocol running on Bluetooth or very few depending on the function of the device. The Protocol field lists the Bluetooth protocol that is being run on the Bluetooth device.

Capabilities

On top of the ports and protocols that are running on a device, there is usually some kind of application that takes user input. That user input can also be attacked to compromise the device. This section is for general higher layer capabilities in software. The table is as follows:

ID	Name	Data Input	Dependency	Description	C

The Data Input is the type of data that the user is passing through to the device. These could be values, characters, or even radio buttons.

Miscellaneous

There are many ways to communicate between devices and newer ways in the future. No standardization could encompass every type of communication without becoming too complex or too redundant. Communications not fitting in any of the prior sections would go into this section. Covert channels might often be documented here. The table is as follows:

ID	Name	Channel	Data Link	Dependency	Description	C

The Channel field is the physical layer of communication that sending or receiving data. This could include temperature or humidity, anything not yet covered by the previous sections that the device can sense or broadcast data. The Protocol field would be the data link layer that specifies how the data is interpreted. An IEEE standard can be written in the field or a 'Custom' protocol can be described in the description.

3) Evaluation and Results

There exist many different technologies on the market. Describing each device on the market and in the future while also being flexible is an integral part of being an ASDL.

3.1) Properties

There are a few properties that ASDL needs to satisfy. These properties are:

- Ability to show interdependent structures
- Extensible to newer technologies
- Allow for uncommon technologies
- Allow for unknown fields

The ability to show interdependent structures is important for attack surfaces as compromising one technology can show a link in which another technology can be compromised. Extensibility allows for ASDL to remain relevant and useful for a long time. There are many technologies and accounting for all of them individually is infeasible. ASDL adds a miscellaneous section such that

these technologies are not excluded, especially if something depends on them. Unknown fields could be theories/guesses on how a device works. This allows for documentation to continue despite not knowing what might lie underneath a device's surface.

3.2) Goals

For ASDL to be accepted, it needs to hit a few goals. These goals are:

- Concise and easy to use
- Flexible for less effort

By setting certain fields and norms for each field, the required amount of writing for each field is less and can be expressed in less words. This would reduce the amount of redundant information being convey in each row. Flexibility really helps reduce the effort as not every row has to be specified and can be combined. For example, combining TCP/IP into one row and specifying multiple ports such that the writer does not have to specify each protocol for each port independently.

3.3) Examples

To further show how ASDL functions, a few examples of documenting attack surfaces with ASDL are shown below.

A simple example is a TP-Link smart plug. The model being documented is the Mini HS105.

HID/Displays

ID	Name	Interaction	Dependency	Description	C
1	Power Button	Button Press		Pressing the button toggles the power on or off	X

Physical Ports

ID	Name	Standardization	Dependency	Description	C
2	Input	Plug Type B		Takes in power for connected device and smart plug	X
3	Output	Plug Type B		Outputs power to connected device	X

Wireless Interfaces

ID	Name	Frequency Range	Dependency	Description	C
4	Wi-Fi Card	2.4-2.48 GHz			X

Wi-Fi/Network/Mobile

ID	Name	Port(s)	Protocol	Dependency	Description	C
5	TCP/IP	9999	TCP/IP	4		X
6	Command Port	9999	Custom	5	The smart plug takes commands to the port from the network	X

Bluetooth

ID	Name	Protocol	Dependency	Description	C
7	Kasa App Pairing	L2CAP+	4	Unknown exact protocol, but initial pairing of IoT device is suspected to be Bluetooth before Wi-Fi connection is set up	

Note that not all the sections are included. If a section has no items in it, that section is excluded entirely to simplify ASDL. The protocol for the Kasa app pairing is unknown here and could be found to be different or more in depth. ASDL does not have to be precise or extremely detailed but can be if it needs to. The inclusion of protocol stacks (for example TCP/IP) could help with detailing ports but does not need to be included as such a protocol would be assumed with the higher layer protocols.

Another example is some Skullcandy Bluetooth headphones. The model being documented is the SESH S2TDW.

HIDs/Displays

ID	Name	Interaction	Dependency	Description	C
1	Left Earbud LED	Colored Light	7	Displays the status of the earbud	X
2	Right Earbud LED	Colored Light	7	Displays the status of the earbud	X
3	Left Earbud Button	Press	7	Controls the audio sent to earbuds and power	X

4	Right Earbud Button	Press	7	Controls the audio sent to earbuds and power	X

Sensors/Actuators

ID	Name	Function	Dependency	Description	C
5	Left Earbud Speaker	Sound	7		X
6	Right Earbud Speaker	Sound	7		X

Wireless Interfaces

ID	Name	Frequency Range	Dependency	Description	C
7	Bluetooth card	2.4 GHz		Connects the phone to the Left Earbud. Exists only in the Left Earbud	X
8	Wireless Chip	Unknown		Connects the Right Earbud.	X

Bluetooth

ID	Name	Protocol	Dependency	Description	C
9	Audio	Audio	7		X

In this example we have some interchanging dependencies. The LEDs are dependent on the Bluetooth audio connection to show different statuses. However, the speakers are also dependent on the audio connection and can function without. The difference between the two is that the main functionality of the LED is used whether the Bluetooth connection is existent or not while the speakers do not implement their main functionality unless the Bluetooth connection is made. As such the LED can function fully without the connection and therefore is not completely dependent on the Bluetooth audio.

The last example is a wireless router. The model being documented is the Netgear Wireless-G Router WGR614 v9.

HIDs/Displays

ID	Name	Interaction	Dependency	Description	C

1	LED Lights	Colored Light	3,7	Shows the status of the router and connected ports	X

Physical Ports

ID	Name	Standardization	Dependency	Description	C
2	Power	AC adapter			X
3	Ethernet	Ethernet		There are 5 ethernet ports on the device with one being designated for internet	X

Wireless Interfaces

ID	Name	Frequency Range	Dependency	Description	C
4	Wireless Card	2.4 GHz			

Wi-Fi/Network/Mobile

ID	Name	Port(s)	Protocol	Dependency	Description	C
5	Router Setup	80	HTTP	3,4	Accessed using the domain name www.routerlogin.net	X
6	Packet Transport		IP	3,4		X

Capabilities

ID	Name	Data Input	Dependency	Description	C
7	Links	Mouse Click	5	The site is navigated by clicking on links and buttons	X
8	Text boxes	Characters	5	There are a few text boxes that require character input such as the password change	X
9	Radio buttons and drop boxes	Mouse Click	5	Some fields can only be changed using the radio buttons and drop boxes	X

Miscellaneous

ID	Name	Channel	Data Link	Dependency	Description	C

10	Router Temp.	Heat	Linear correlation		A possible covert channel with the router getting hotter with heavier use	

The capabilities field is very minimal and does not completely outline every feature in the router's webpage. However, webpages can have a lot of different inputs and fields, even with the simplest of webpages. The best way to simplify this is to group by types of data input, which makes things a lot easier. On the other hand, capabilities of a website can be written one by one for each feature. It is ultimately up to the writer how detailed they want this field to be.

4) Conclusion

In conclusion, ASDL can be used for the representation of electronic communication for any electronic device. The XML format of ASDL organizes the data for use by both people and computers, allowing for the ability of automation. ASDL also shows dependencies and can trace any compromises in one technology to another technology. Newer technologies can also fit in the categories or in the miscellaneous field. If more detail is needed, it can be added in either multiple rows or elaboration in the description field. In these ways, ASDL satisfies the basic requirements of describing electronic communications.

However, satisfying basic requirements is never good enough. ASDL has to be user-friendly for both the writers and readers. By splitting various components of communication, navigation of communications becomes easier. Most fields are also optional or can be compressed such as to not require too much work on the writer's behalf.

Future Works

At the time of writing, standardizing the documentation of attack surfaces on a device is not a well-researched topic. As such, there is a lot of work yet to be done. As technologies grow and change, some categories in ASDL might become unused or often irrelevant. ASDL might have to be changed and updated from time to time to accommodate for technological growth.

Even though XML is relatively user-friendly, it can still become encumbering with lots of data entries. A graphing software that maps and graphs ASDL would alleviate some of the encumbrance and provide a unique way of looking at the data.

5) References

[1] Daniels, Thomas & Kuperman, Benjamin & Spafford, Eugene. (2000). Penetration Analysis of a XEROX Docucenter DC 230ST: Assessing the Security of a Multi-purpose Office Machine.

[2] G. A. Olmstead, B. Kilic, D. Yum, K. Leung, A. Sharma, Y. Zhang, "System and method for electronic communications." U.S. Patent 10,032,450 B2, issued July 24, 2018.

[3] D. W. Malik, “System and method for sorting electronic communications.” U.S. Patent 7,930,352 B2, issued April 19, 2011.
[4] Manadhata, Pratyusa & Wing, Jeanette M. Measuring a System’s Attack Surface. http://www.cs.cmu.edu/~wing/publications/ManadhataWing04.pdf