Explainable Security

Luca Viganò
Department of Informatics
King's College London, London, UK
luca.vigano@kcl.ac.uk

Daniele Magazzeni
Department of Informatics
King's College London, London, UK
daniele.magazzeni@kcl.ac.uk

***Abstract*—In 2017, the Defense Advanced Research Projects Agency (DARPA) launched the Explainable Artificial Intelligence (XAI) program that aims to create a suite of new AI techniques that enable end users to understand, appropriately trust, and effectively manage the emerging generation of AI systems. In this paper, inspired by DARPA's XAI program, we propose a new paradigm in security research: Explainable Security (XSec). We discuss the "Six Ws" of XSec (Who? What? Where? When? Why? and How?) and argue that XSec has unique and complex characteristics: XSec involves several different stakeholders (i.e., the system's developers, analysts, users and attackers) and is multi-faceted by nature (as it requires reasoning about system model, threat model and properties of security, privacy and trust as well as concrete attacks, vulnerabilities and countermeasures). We define a roadmap for XSec that identifies several possible research directions.**

***Index Terms*—Security paradigm, Security model, Threat model, Security properties, Privacy, Trust, Usable security, Security economics, Gamification**

1. Introduction

The security of information, data, processes, software, protocols, computers, networks and systems is notoriously a challenging problem (and very often an undecidable one). Security is difficult. It is difficult to achieve, to reason about, to apply, to understand, to teach. It is difficult to *explain*.

In 2017, the Defense Advanced Research Projects Agency (DARPA) launched the *Explainable Artificial Intelligence (XAI)* program that aims to create a suite of new AI techniques that enable end users to understand, appropriately trust, and effectively manage the emerging generation of AI systems. Some research on explainable AI had already been published before DARPA's program (e.g., [45], [48], [1], [2], [7], [44], [42], [32], [19]), but XAI encouraged a large number of researchers to take up this challenge. In the last couple of years, several publications have appeared that investigate how to explain the different areas of AI, such as *machine learning* [21], *recommender systems* [30], *robotics* and *autonomous systems* [35], [43], [20], *constraint reasoning* [18] and *planning* [10], [17], [24].[1]

1. See also [28] and the other papers listed at http://home.earthlink.net/~dwaha/research/meetings/faim18-xai/.

In this paper, inspired by the XAI program, we propose a new paradigm in security research:

Explainable Security (XSec).

Some pioneering works on explaining security have focused on *security for relational databases* [6] and on *explanation and trust* [34]. In [6], Bender, Kot and Gehrke propose a new model in which policy decisions are explainable. In this model, instead of simply rejecting an unauthorized query by a principal, the system provides the principal with a concise explanation of why the query was rejected and what additional permissions the principal would need to be granted for a successful execution. The principal can then refine the query or request additional permissions based on the explanation provided.

In [34], Pieters investigates the relation between explanation and trust, focusing in particular on expert systems and e-voting systems. Pieters observes that

> *In artificial intelligence, explanations are usually provided by the system itself. In information security, explanations are provided by the designers. Nonetheless, in* both *artificial intelligence and information security, the role of explanations consists for a major part of acquiring and maintaining the trust of the user of the system.*

He discusses how explanations are required for trust:

> *Here, the question is how it is possible to communicate the analysis that experts have made of a security-sensitive system to the public. Why is it secure? Or, more appropriately: How is it secure?*

Explanations are thus

> *thought to bridge the gap between 'actual security' and 'perceived security'.*

Pieters also discusses two main goals that an explanation may have: *transparency* (e.g., to allow users to understand what the designers have done to protect them) and *justification* (e.g., offering reasons for an action). He contrasts *explanation-for-trust* (i.e., explanation of how a system works, by revealing details of its *internal* operations) with *explanation-for-confidence* (i.e., explanation to make the user feel comfortable in using the system, by providing information on its *external* communications).

We argue that XSec is a difficult problem and that it has unique and complex characteristics. In fact, XSec is more complex than what is discussed by Bender et al. and by Pieters; it is also tightly connected to, but different

from, *usable security* [33], [47], *security awareness* [5], [49] and *security economics* [3]. This is because XSec involves several different stakeholders (i.e., the system's developers, analysts, users and attackers) and is multi-faceted by nature (as it requires reasoning about system model, threat model and properties of security, privacy and trust as well as about concrete attacks, vulnerabilities and countermeasures).[2]

XSec is thus an exciting novel paradigm that requires a full-fledged and heterogeneous research program to be realized. In the following, we define a roadmap that identifies several possible research directions. To describe the challenges of XSec and how they could be tackled, we proceed by discussing the "Six Ws" of XSec summarized in Figure 1: Who? What? Where? When? Why? and How?

2. Who?

Consider a generic *system*, where we use here "system" to refer to a generic process, software, protocol, computer, network, cyber-physical system, critical infrastructure, etc. that processes information/data whose security must be protected, where "security" similarly generically refers to one or more of the security properties of interest, including confidentiality, integrity, availability, authentication, authorization, non-repudiation, accountability, unobservability, privacy, etc.

The *dramatis personae* of XSec are:

- the *designer* (and/or *developer*) of the system, who has designed (and/or developed) the system to guarantee a number of specified security properties;
- the *user* (and/or *client*) of the system, who can typically be assumed to be an honest non-expert who might commit mistakes that make the system vulnerable;
- the *attacker* (or more *attackers*) of the system, who searches for and exploits vulnerabilities of the system for reasons or profit, fame, reward, etc.;
- the *analyst* of the system, who carries out a semi-formal or formal analysis of the system at design time (or based on the specification of a deployed system) or tests the system at runtime (e.g., using penetration testing, vulnerability-based testing or model-based testing);
- the *defender* of the system, who attempts to protect the system, e.g., by monitoring the activities of the system and reacting to the attacker's actions.

For some situations, the recipient of the explanation will be an agent rather than a human, and we can then contrast *internal explanations* (designed for software agents) and *external explanations* (designed for humans, which is what XAI research typically focuses on).

2. One could argue, for instance, that in order to be usable security needs first to be explainable, or vice versa. However, there is no shortage of real-life security systems and solutions that are usable without being explainable or without explaining themselves. Think about the most recent smartphones, which come without any user manual. They are usable but provide little explanation so that often users don't understand why and how they work, with possible dire consequences for the ultimate security of the data that the smartphone will manage. Similarly, explanations will need to be understandable by the human users to contribute to usability.

Some of the above roles might actually be played by the same "principal" (agent or human), as the designer might for instance act also as analyst or defender, the analyst might also provide immediate defense and the attacker might be a user of the system. The literature is full of examples of vulnerabilities caused by mistakes by designers or users, along with details of the corresponding attacks. Some of these attacks could have been prevented by better explanations. In fact,

> all of these roles might require explanations or need to act as explainer.

For instance,

- a designer and/or an analyst might need to explain to the user how to interact with the system, why the system is secure and why it carries out a particular action (in line with what is discussed in [6], [34]);
- a user or client might need to explain to the designer or the analyst how he expects the system to behave and how they typically interact with the system, to allow the designer to elicit the requirements for building the system in the first place and to allow the analyst to validate the security of user interactions;
- an attacker might need to explain the attack strategy to his accomplices so that they can attack in coordination, or he might have used a complex penetration testing tool to test the system for vulnerabilities and now needs the tool to explain to him the attack trace (or attack plan or strategy) that has been identified so that he can carry out the attack for real;
- an analyst might need to explain to the designer how to improve the system's security or to the defender how and what to defend;
- a defender might similarly need to understand possible attack traces in order to take action against them as well as explain to the users how they should behave to protect the system and themselves.

In addition to this, it is also necessary to tackle the research question of what actually constitutes a good (and secure) explanation, as we discuss in more detail in the following sections.

Before we do so, let us consider a concrete example that arises from the observation that when dealing with sensitive data, classical authentication solutions based on username-password pairs are not enough.

Example 2.1. The "General Data Protection Regulation" [15] mandates that specific security measures must be implemented, including *multi-factor authentication (MFA)*, an authentication solution that aims to augment the security of the basic username-password authentication by exploiting two or more authentication factors (see, e.g., [40], [41]). In [14], MFA is for instance defined as:

> *"a procedure based on the use of two or more of the following elements — categorised as knowledge, ownership and inherence: i) something only the user knows, e.g., static password, code, personal identification number; ii) something*

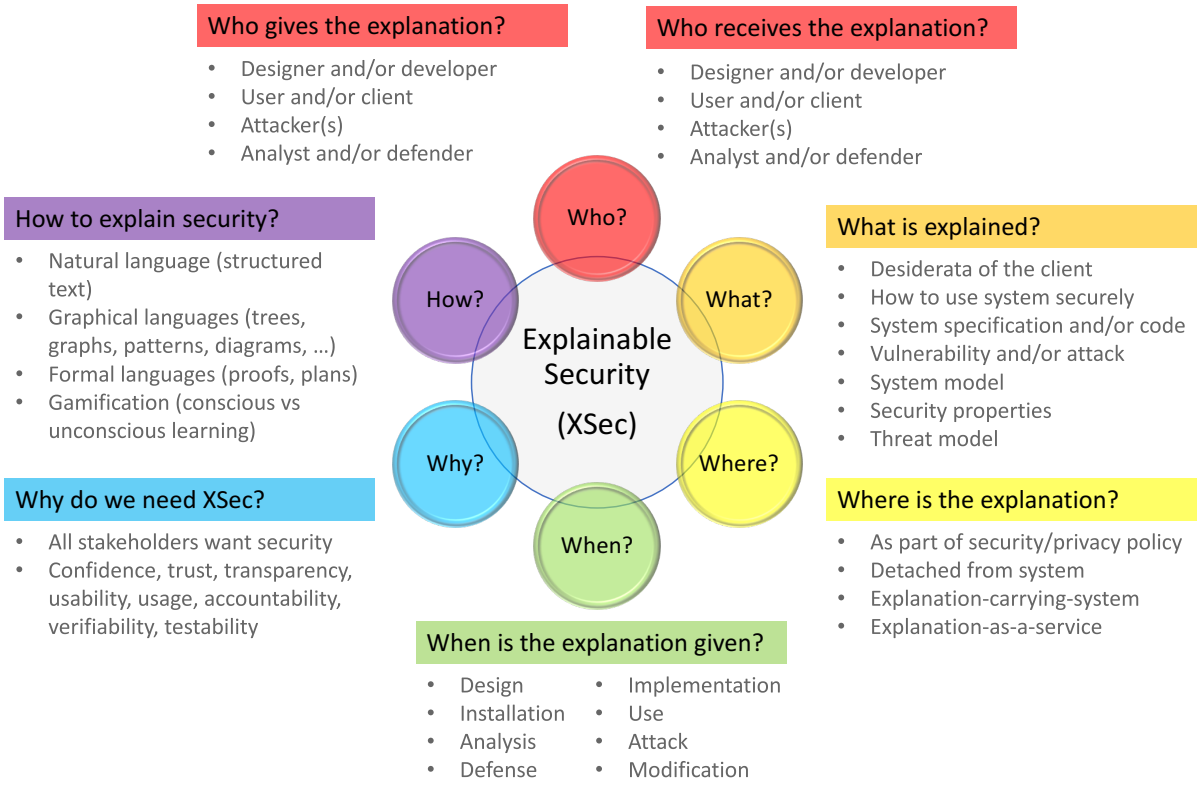

Figure 1: The Six Ws of Explainable Security

only the user possesses, e.g., token, smart card, mobile phone; iii) something the user is, e.g. a biometric characteristic, such as a fingerprint. In addition, the elements selected must be mutually independent [...] at least one of the elements should be non-reusable and non-replicable".

The underlying idea is that the more factors are used during the authentication process, the more confidence a service has that the user is correctly identified. This is the basic explanation provided by the designer to the user to justify a more involved authentication that the user might perceive as cumbersome. However, the user might also need to be told that choosing a weak password is a bad idea even in the case of MFA. Two attacker accomplices who carry out a coordinated attack against the two components of MFA might need to explain their sub-attack to each other to ensure their ultimate success. The analyst who has discovered an attack to the MFA system might need to explain to the designer why the attack succeeded and how to patch it. The analyst/defender might also need to explain to the users why they should, e.g., abandon the use of one of the elements they had been using so far and switch to using another pair of elements; for instance, because the new password that a user has chosen is too weak and thus easily guessable, or because the device on which the user is trying to authenticate does not include a biometric reader. ■

There are thus many things to explain by/to many different stakeholders, which is one of the main reasons why XSec, even in the case of a relatively simple example such as MFA, is a challenging endeavor. In the following, we discuss the five remaining Ws, although we have already anticipated much of the discussion here (which is somewhat unavoidable given that the Ws are not independent but quite deeply intertwined).

3. What?

It is not enough to explain the system in a generic way. First of all, the different stakeholders will need explanations at different levels of detail and with different aims:

- designers/developers will need an explanation of the desiderata of the client that is detailed enough for them to be able to realize the system in a satisfactory and secure way (e.g., if the client wants a system that replaces passwords with face recognition);
- non-expert users will need an explanation that increases their confidence and trust, and that also teaches them how to use the system correctly and securely (e.g., if passcodes are used as back-up in case face recognition fails as in the iPhone X, then the user should be made aware that the passcode ought to be strong enough and not guessable such as a date of birth or a phone number, since otherwise an attacker who steals the iPhone X will obviously fail face recognition but the iPhone X will allow him to get access by guessing the passcode);
- analysts will need access to the system's specification or to the system's code in order to be able to create a model to analyze or to be able to generate and execute test cases;

- designers/developers/defenders will need an explanation of a vulnerability and related attacks in order to implement patches or defenses;
- attackers will need an explanation of how to exploit possible vulnerabilities, of why their attacks failed and of the implications of new security techniques on their attack strategies.

Second, several different "things" will need to be explained, including:

- the *system* and the *system model* used for design, implementation and analysis, e.g., the model of how MFA actually works;
- the *security properties* that the system should guarantee, e.g., the authentication provided by MFA can be used as a basis to provide authorization, integrity, confidentiality, non-repudiation and so on;
- the *threat model* that has been considered by designers, developers and analysts, highlighting, in particular, the fact that a system might be secure against one threat model but insecure against another (e.g., a system might be secure against an outside attacker but insecure against insider attacks) or the fact that the successful MFA of a user won't prevent the system from being attacked when that user turns out to be malignant and reveals, say, trade secrets of the company he works for;
- the actual *vulnerability* and related *attack* that has been discovered will need to be explained to the attacker (especially when the attacker used a tool to search for an attack and now needs to carry out it concretely), along with the costs and benefits of the attack;
- the possible *countermeasures* for the discovered *vulnerability* and related *attack* will need to be explained (along with the vulnerability and the attack) to the honest stakeholders who will need to understand the attack's impact, its risk and the mitigation strategies.

To that end, it will be helpful to answer a number of questions, including:

- What is actually secure? Which parts of the system? Which security properties are guaranteed and for how long? (For instance, authentication is typically granted for a session, which expires after some amount of time.) Which features are insecure and why and how can they be attacked? Are there different levels of security (e.g., for users with different rights)?
- What is the threat model considered? Does it include insiders and outsiders? Who are the potential attackers and what do they want? Why do they want it?
- How does the attack look like and how "difficult" is it? How expensive is it? (It does not make sense to use a one million dollar machine to mount an attack with a loot of a few thousand dollars.) How long will the attack take? (If students try to steal the questions of their next exam but their attack takes so long that they get hold of the questions only after the exam has been given by the professors, then there is actually no point.) This requires reasoning quantitatively about the economics of the attack (including costs, performance, time) but also about the trade-off between attacking and the risk of being discovered.
- Under which assumptions and conditions is the system assumed to be operating securely or has been proved to be secure? For instance, many security analyses (e.g., of protocols or web applications) typically assume a *Dolev-Yao-style attacker* [13] who controls the network but cannot break cryptography, which is quite a strong assumption to make as cryptography might indeed be broken (by classical computers and even more so by quantum computers if and when they will be realized in their full capacity); on the other hand, relaxing this assumption and considering an attacker who might be able to break cryptography typically complicates the analysis (the problem is undecidable anyway) and the ability of an analyst to prove security guarantees.
- What are the *legal implications* of the explanation? Is the explanation "binding"? This would require, for instance, explaining how the system works and what is expected of the user, possibly including a digital signature to acknowledge the receipt and understanding of the explanation. In case of an attack, this would also require explaining what happened and why, and what countermeasures can be taken (and by whom).

4. Where?

We have already addressed the question of which "parts" of the system need to be explained in the "What?" section. Now we focus briefly on the question of where the explanations should be made available. A number of different options are available here, including the following four main ones:

- One could include the explanations to the users as part of the security/privacy policy, but it is well known that users typically ignore the policy and scroll down as quickly as possible so that they can get on with their interaction with the system.
- One could completely detach the explanation from the system, e.g., by making it available on a different webpage, but it is unclear to us if and how the relevant stakeholders will be made aware of where to find the explanation and whether they will decide to trust it.
- One could consider, similar to [9], a sort of *explainable security as a service*, where stakeholders interact with an expert system to obtain and/or provide explanations.
- One could proceed in the style of *proof-carrying code* [31], "appending" a possibly digitally-signed explanation to the system to achieve a *security-explanation-carrying-system*. We believe that this is the most promising direction, but it will of course require considerable work to protect the explanation from attacks and actually explain to

the stakeholder how they can access it and make use of it.

5. When?

We want Explainable Security and we want it now! Jokes aside, the many vulnerabilities that are reported daily, including some of our most widespread and supposedly secure systems (consider, e.g., recent attacks against: TLS; PGP; processors; dropbox, one drive, iCloud and other cloud systems; biometric authentication systems; e-commerce and e-banking systems; e-voting systems, etc.), are witness to the fact that security is indeed difficult to achieve (which is why security has been and still is one of the hottest research topics) but also that in many cases security systems are difficult to explain to the different stakeholders.

We need to explain security when the system is

- designed,
- implemented,
- deployed and installed,
- used,
- analyzed,
- attacked,
- defended,
- modified,

and possibly even when the system is decommissioned and replaced, so that the different stakeholders understand why this decision was taken and how the new system will improve over the old one.

In particular, explanations will need to be defined and provided at *design time* (when the system is developed) but also at *runtime* (when the system is running). For the runtime case, think, e.g., of a critical system, critical infrastructure or cyber-physical system such as a nuclear power plant in which a supervisor is in charge of setting high/low security levels and of intervening in the case of an ongoing attack to estimate the success chances of the attack, understand its impact on the system and adopt possible countermeasures (see, e.g., [26], [25]). The attack could have disastrous consequences (e.g., manipulating the SCADA and PLC systems of a power plant as the Stuxnet Worm [16]) or (appear to) be non-threatening as it manipulates sensors and actuators of the system but without bringing them outside of their tolerance zone so that the supervisor actually decides not to intervene.

6. Why?

This is easiest question to answer: because all the different stakeholders of a system want it to be secure (well, with the exception of the attacker, of course). Explanations will help increase confidence, trust, transparency, usability and concrete usage (in the sense that users will be more keen to adopt the system), accountability, verifiability and testability.

7. How?

As we already remarked above, the different stakeholders will need explanations at different levels of detail and with different aims, and these explanations will need to be comprehensible, timely and accurate (among other properties). The explanations will need to be written in a language (and with a description strategy) suitable for the intended audience, including[3]

- *natural language* (used to produce informal but possibly structured text written in English or any other language understandable by the audience);
- *graphical languages* such as explanation trees, attack trees, attack-defense trees, attack graphs, attack patterns, message-sequence charts, use case diagrams ...;
- *formal languages* including proofs and plans;
- games that have been produced as the result of a *gamification* process to teach users how to interact with a system (although one could actually object that such games often provide for some "unconscious learning" in which the user learns how to interact but without really understanding why).

It should also be investigated whether one learns more by seeing a proof of the security of the system or by being shown an attack against an insecure system. Both are of course useful, but they explain different things in a different way, and they can both be traced back to the question asked by Pieters [34] of "how it is possible to communicate the analysis that experts have made of a security-sensitive system to the public".

It will also be necessary to evaluate security explanations originating from applications of XAI and other areas of computer science, and possibly even social sciences, psychology and other disciplines. Careful subject studies will need to be designed to assess measurement categories such as: a priori measures of explanation quality, user satisfaction, mental model understanding, and user-machine task performance.

Moreover, the explanation processes will themselves have to be designed properly, tested thoroughly and deployed correctly, and it will be useful to investigate the trade-off between such explanation processes and the security threats.

We expect that many of the How? questions posed in the specific case of security will actually be answerable by suitably adapting and extending the techniques and tools that have been and are being developed for XAI. Still, we conclude the discussion of the Ws by considering again the claim that we made above that XSec has unique and complex characteristics, and is more challenging than the pioneering research on explanations in security [6], [34]. Let us illustrate this by an example that shows that XSec calls for proper extensions of the research on XAI and for novel investigations.

In Explainable Planning [8], [17], [23], one of the questions the planner should answer is why things cannot be done and why and how one needs to replan. Similarly, in the model of Bender et al., the relational database system provides the principal with a concise explanation of why the query was rejected and what additional

3. *Popular films and artworks* can also be used to explain security [46], in the spirit of the growing body of literature on using film to explain and teach different disciplines such as (to name only a few) philosophy [4], [29], history [27], social sciences [37], management and organizational behavior [11], [12], mental illness [36].

permissions the principal would need to be granted for a successful execution. If one considers a more general security system, however, such an explanation might make the system less secure! This is because the explanation itself might reveal security-sensitive information.

For instance, the attacker might not know whether a certain person is indeed a user of the system: trying to login pretending to be that user and being told that the user does not exist, or that the password is wrong, or that the user needs more privileges to be able to carry out some specific action already constitutes a leak of information.[4] Hence, explanations need to be "relativized" and in some cases made less "powerful" by withholding certain details. But a less powerful explanation is essentially an incomplete explanation, which will be ignored or not fully achieve its purpose. The quest for a reasonable trade-off thus makes XSec particularly challenging.

8. A more detailed example

Let us now return to Example 2.1 and expand the discussion of the explanations that can be considered in the case of multi-factor authentication processes and protocols.

Example 8.1. We have already seen a number of explanations that could be useful for MFA. Let us consider these explanations again, providing more details when possible, and add a few interesting cases.

- The designers/developers could explain to the users that using two or more factors during the authentication process yields a stronger authentication. In fact, this will more likely be split into two different explanations, the one that the designers/developers give to their client, say an online bank, which in turn will provide the authentication service to the end-user and explain to them how to use it.
  The explanation by the designers/developers might also include a description of the hardware and software needs, e.g., suggesting to the bank whether they should provide dedicated authentication devices to the end-users or whether these could use their smartphones. This explanation will be provided before the system is deployed and could be part of the development contract. It will likely include diagrams to describe usage but also proofs or semi-formal justifications of the degree of security achieved.
  The explanation by the bank to the end-users might also include an explanation of why MFA is more secure than "standard" authentication and on what kinds of smartphones and operating systems it works. This will be provided before the user authenticates for the first time (or possibly be suggested as an alternative every time the user logs in using standard authentication) and it could be part of the Terms and Conditions of the online banking contract. This could also provide the opportunity to explain to the user that choosing a weak password is a bad idea even in the case of MFA.
- Two attacker accomplices who carry out a coordinated attack against the two components of MFA might need to explain their sub-attack to each other to ensure their ultimate success. Indeed, as early as 2005, Schneier [38] warned about the pitfalls of MFA and how attackers have been modifying their tactics and developed attacks to MFA based on man-in-the-middle and trojans.
  The analyst who has discovered an attack to the MFA system might need to explain to the designer why the attack succeeded and how to patch it. The analyst/defender might also need to explain to the users why they should, e.g., abandon the use of one of the elements they had been using so far and switch to using another pair of elements; for instance, because the new password that a user has chosen is too weak and thus easily guessable, or because the device on which the user is trying to authenticate does not include a biometric reader. Users will also need to be explained the role that they can play in the attacks, e.g., how an attacker using a man-in-the-middle attack is happy to have the user deal with the SMS portion of the login, since he can't do it himself, and how a Trojan attacker relies on the user to log in [38]. Man-in-the-Browser attacks are also possible [39].
- Finally, judging by the confusion that appears to be afflicting users, clients and even some designers/developers of several systems, explanations would be needed to differentiate more clearly between two-factor authentication and *two-step authentication*, where, e.g., a user is asked first to provide a password and then two or more letters from a memorable word.

More explanation scenarios useful in the case of MFA services and protocols can be quite straightforwardly identified. Ultimately, even this simple example serves to show that, as we already discussed above, all stakeholders are likely to be involved in the explanations, both as explanation providers and as explanation receivers, and that explanation will need to provided for different reasons, at different times, in different places and in different guises. ■

9. Conclusions

We see XSec as a new paradigm that brings together many hot topics and current trends but also indicates the need of exploring unchartered territory. This paper has just skimmed the surface of XSec by pointing out some of the main objectives and challenges, and defining a roadmap that identifies several possible research directions.

We have already begun a more detailed investigation of the different techniques and tools that can be exploited or need to be developed to answer the questions posed by the different Ws, as well as formalizing the relationships and interdependences between the Ws. More specifically, we expect that, using the growing amount of results on

4. As another example, consider argument passing within GET protocols in HTTP requests, which basically allow the attacker to understand which is the name of the server's variable that handles the request, thus facilitating the attack. POST protocols are to be preferred, but they can't be used always.

XAI together with the pioneering research on explanations in security and trust as a stepping stone, we will soon be able to provide some concrete answers. To that end, we will also take inspiration from research that have attempted to "bridge the gap" between different research questions and different communities, such as the work of Hoffman on AI planning for penetration testing [22]. We also plan to explore in more detail the connections and synergies between XSec and formal methods, argumentation and planning for security, as we well as with usable security, security awareness and security economics.

This is the beginning of a beautiful friendship between explanations and security, and we plan to be part of the fellowship that nurtures this friendship.

Acknowledgment

A preliminary version of this paper was presented as a poster at the "IJCAI/ECAI 2018 Workshop on Explainable Artificial Intelligence (XAI)". We thank David W. Aha, Fabio Mercorio, Luc Moreau, Diego Sempreboni and the anonymous reviewers of the XAI 2018 workshop and of the HotSpot 2020 workshop for their useful comments and suggestions.

References

[1] R. Aldeco-Pérez and L. Moreau. Provenance-Based Auditing of Private Data Use. In *Visions of Computer Science*, pages 141–152. British Computer Society, 2008.

[2] R. Aldeco-Pérez and L. Moreau. A Provenance-Based Compliance Framework. In *Future Internet – FIS 2010*, LNCS 6369, pages 128–137. Springer, 2010.

[3] R. Anderson. Economics and Security Resource Page, 2018. http://www.cl.cam.ac.uk/~rja14/econsec.html.

[4] T. Ariemma. *La filosofia spiegata con le serie TV*. Mondadori, 2017.

[5] J. M. Banfield. *A Study of Information Security Awareness Program Effectiveness in Predicting End-User Security Behavior*. PhD thesis, Eastern Michigan University, 2016.

[6] G. Bender, Ł. Kot, and J. Gehrke. Explainable Security for Relational Databases. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, pages 1411—-1422. ACM, 2014.

[7] J. Bidot, S. Biundo, T. Heinroth, W. Minker, F. Nothdurft, and B. Schattenberg. Verbal Plan Explanations for Hybrid Planning. In *Proceedings of Multikonferenz Wirtschaftsinformatik (MKWI)*, pages 2309–2320. Universitätsverlag Göttingen, 2010.

[8] R. Borgo, M. Cashmore, and D. Magazzeni. Towards Providing Explanations for AI Planner Decisions. *CoRR*, abs/1810.06338, 2018.

[9] M. Cashmore, A. Collins, B. Krarup, S. Krivic, D. Magazzeni, and D. E. Smith. Towards Explainable AI Planning as a Service. *CoRR*, abs/1908.05059, 2019.

[10] T. Chakraborti, S. Sreedharan, Y. Zhang, and S. Kambhampati. Plan Explanations as Model Reconciliation: Moving Beyond *Explanation as Soliloquy*. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*. IJCAI, 2017.

[11] J. E. Champoux. *Management: Using Film to Visualize Principles and Practice*. South-Western, 2000.

[12] J. E. Champoux. *Organizational Behavior: Using Film to Visualize Principles and Practices*. South-Western, 2000.

[13] D. Dolev and A. Yao. On the Security of Public-Key Protocols. *IEEE Transactions on Information Theory*, 29(2):198–208, 1983.

[14] European Central Bank. Final guidelines on the security of Internet payments, 2014.

[15] European Commission. Regulation EU 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016.

[16] N. Falliere, L. Murchu, and E. Chien. W32.Stuxnet Dossier, 2011.

[17] M. Fox, D. Long, and D. Magazzeni. Explainable Planning. In *Proceedings of IJCAI Workshop on Explainable Planning*, 2017.

[18] E. Freuder. Explaining ourselves: Human-aware constraint reasoning. In *AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 4858–4862. AAAI Press, 2017.

[19] F. Gedikli, D. Jannach, and M. Ge. How Should I Explain? A Comparison of Different Explanation Types for Recommender Systems. *International Journal of Human–Computer Studies*, 72(4):367–382, 2014.

[20] B. Hayes and J. A. Shah. Improving Robot Controller Transparency Through Autonomous Policy Explanation. In *HRI'17: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 303–312. ACM, 2017.

[21] L. A. Hendricks, Z. Akata, M. Rohrbach, J. Donahue, B. Schiele, and T. Darrell. Generating Visual Explanations. In *ECCV 2016: Computer Vision*, LNCS 9908, pages 3–19. Springer, 2016.

[22] J. Hoffmann. Simulated Penetration Testing: From "Dijkstra" to "Turing Test++". In *ICAPS'15: Proceedings of the Twenty-Fifth International Conference on International Conference on Automated Planning and Scheduling*, pages 364–372. ACM, 2015.

[23] J. Hoffmann and D. Magazzeni. Explainable AI Planning (XAIP): Overview and the Case of Contrastive Explanation (Extended Abstract). In *Reasoning Web. Explainable Artificial Intelligence - 15th International Summer School 2019, Bolzano, Italy, September 20-24, 2019, Tutorial Lectures*, LNCS 11810, pages 277–282. Springer, 2019.

[24] P. Langley, B. Meadows, M. Sridharan, and D. Choi. Explainable Agency for Intelligent Autonomous Systems. In *AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 4762–4763. AAAI Press, 2017.

[25] R. Lanotte, M. Merro, A. Munteanu, and L. Viganò. A Formal Approach to Physics-based Attacks in Cyber-Physical Systems. *ACM Trans. Priv. Secur.*, 23(1):3:1–3:41, 2020.

[26] R. Lanotte, M. Merro, R. Muradore, and L. Viganò. A Formal Approach to Cyber-Physical Attacks. In *Proceedings of CSF*, pages 436–450. IEEE, 2017.

[27] A. S. Marcus, S. A. Metzger, R. J. Paxton, and J. D. Stoddard, editors. *Teaching History With Film: Strategies for Secondary Social Studies*. Routledge, 2 edition, 2018.

[28] T. Miller. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artif. Intell.*, 267:1–38, 2019.

[29] R. Mordacci, editor. *Come fare filosofia con i film*. Carocci editore, 2017.

[30] K. Muhammad, A. Lawlor, and B. Smyth. Explanation-based Ranking in Opinionated Recommender Systems. In *Proceedings of the 24th Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*. CEUR Workshop Proceedings, 2016.

[31] G. C. Necula. Proof-Carrying Code. In *POPL'97: Proceedings of the 24th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*, pages 106–119. ACM, 1997.

[32] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos. A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Min Knowl Disc*, 24(3):555–583, 2012.

[33] B. D. Payne and W. K. Edwards. A brief introduction to usable security. *IEEE Internet Computing*, 12(3):13–21, 2008.

[34] W. Pieters. Explanation and trust: What to tell the user in security and AI? *Ethics Inf Technol*, 13:53–64, 2011.

[35] S. Rosenthal, S. P. Selvaraj, and M. M. Veloso. Verbalization: Narration of Autonomous Robot Experience. In *IJCAI'16: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 862–868. IJCAI, 2016.

[36] L. C. Rubin, editor. *Mental Illness in Popular Media: Essays on the Representation of Disorders*. McFarland & Co, 2012.

[37] W. B. Russell III. The Art of Teaching Social Studies with Film. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 85(4):157–164, 2012.

[38] B. Schneier. Two-Factor Authentication: Too Little, Too Late. *Communications of the ACM*, 48(4):136, 2005.

[39] B. Schneier. The Failure of Two-Factor Authentication, 2012. https://www.schneier.com/blog/archives/2012/02/the_failure_of_2.html.

[40] G. Sciarretta, R. Carbone, S. Ranise, and L. Viganò. Design, Formal Specification and Analysis of Multi-Factor Authentication Solutions with a Single Sign-On Experience. In *POST 2018: Principles of Security and Trust*, LNCS 10804, pages 188–213. Springer, 2018.

[41] G. Sciarretta, R. Carbone, S. Ranise, and L. Viganò. Formal Analysis of Mobile Multi-Factor Authentication with Single Sign-On Login. *ACM Transactions on Privacy and Security*, 23(3), 2020.

[42] B. Seegebarth, F. Müller, B. Schattenberg, and S. Biundo. Making Hybrid Plans More Clear to Human Users - A Formal Approach for Generating Sound Explanations. In *ICAPS'12: Proceedings of the Twenty-Second International Conference on International Conference on Automated Planning and Scheduling*, pages 225–233. ACM, 2012.

[43] R. Sheh. "Why did you do that?" Explainable intelligent robots. In *Proceedings of AAAI Workshop on Human-Aware Artificial Intelligence*, 2017.

[44] S. Sohrabi, J. A. Baier, and S. A. McIlraith. Preferred Explanations: Theory and Generation via Planning. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011*, pages 261–267. AAAI Press, 2011.

[45] W. Swartout, C. Paris, and J. Moore. Explanations in knowledge systems: Design for explainable expert systems. *IEEE Expert*, 6(3):58–64, 1991.

[46] L. Viganò. Explaining Cybersecurity with Films and the Arts. In *Imagine Math 7*. Springer Nature, 2020.

[47] R. Wash and M. Zurko. Usable Security. *IEEE Internet Computing*, 21(3):19–21, 2017.

[48] L. Ye and P. Johnson. The impact of explanation facilities on user acceptance of expert systems advice. *MIS Quarterly*, 19(2):157–172, 1995.

[49] E. Yildirim. The importance of information security awareness for the success of business enterprises. In *Advances in Human Factors in Cybersecurity*, Advances in Intelligent Systems and Computing 501, pages 211–222. Springer, 2016.