Eric: Hi, hope you enjoyed last week's episode with Somos and Genobank, where like this week, we really sharpened our focus on encryption and practice. Now I had Kevin McCarthy from Inpher on my podcast on March the 15th, and that was a great episode, and he introduced me to TNO, another independent research organization, whose mission is to connect people and knowledge, to create innovations that boost the competitive strength of industry, and the wellbeing of society in a sustainable way.

Now that is a mission statement, and they also cover a lot of ground. One of which is privacy and privacy- enhancing technologies. They collaborate with the government and academia, as well. So, it really lets them fulfill that mission. That mission is real. When I first started down this road, I thought homomorphic encryption was the big story.

But what I quickly learned was that that is not the case. The big story is about optimizing and integrating a wide array of emerging privacy and learning technologies to better enable outcome. But what does that actually mean? This is what Thomas and I delve into in this episode. So, this episode, like the last one is less about what the future could be and more about what the future is.

Eric: Hi, this is Eric Hess with The Encrypted Economy. Today, I am delighted to have Thomas Attema, a cryptographer at TNO, which is an applied research institute in the Netherlands. I was introduced to Thomas because TNO is a, you know, as I said, an applied research institute, focusing on privacy-enhancing technologies and their applications, and they have really done an awful lot in a space that is caught a lot of attention from a number of cryptographers in this space. So, I thought it was an excellent opportunity to bring somebody on who had a real practical application perspective on the use of privacy-enhancing technologies and has done research and contemplated future use cases.

So, Thomas, with that, welcome to The Encrypted Economy.

Thomas: Thank you. Thanks for having me.

Eric: Excellent. Thomas, I gave a little bit on your, on who you were, but if you could give us maybe a little bit of background on you, and what got you interested in cryptography.

Thomas: Yeah, sure. Okay, so, my background is in mathematics. I studied mathematics in Utrecht, and also in the Netherlands, and I graduated in algebraic number theory back in 2013, and after graduating from a university, I started working at TNO. So, as you already mentioned, TNO is a blind research institute, and one of the topics we work on a lot is privacy-enhancing technologies, but we do much more than that.

So, the, the topics we work on at TNO range from solving all sorts of societal challenges, for example working on climate change, but also security of the digital society. So, all sorts of different topics, and I started working at TNO on topics not related to cryptography at all.

So, I worked on the efficiency of telecommunication networks on structural reliabilities. So basically, the degradation of bridges in the Netherlands but also in cyber security and cruciate detection systems. So, a wide variety of topics, but the common denominator in all these topics was that we were applying mathematical techniques to solve network-related problems in these different domains.

So, I did all sorts of things, very different from cryptography, and in 2016, I got to know one of my colleagues and he told me about his work on multi-party computation and cryptography in general, and he got me really enthusiastic. So, so that is when I decided I wanted to also work on this topic and my background in algebraic number theory really fit as a fit as well or matched well with cryptography.

So, I could jump right into this. This was, this is really cool, and nowadays I work full time on cryptography. I am working on topics as multi-party computation, but also zero-knowledge proofs, and post-quantum cryptography. And, and in 2018 I also started doing a part-time PhD at the cryptology group of CWI, which is a fundamental research institute in the Netherlands on mathematics and computer science.

And I am also affiliated to the University of Leiden with respect to the, this part-time PhD program. So that is a quick introduction.

Eric: So, it seems like you just cannot get enough. Huh?

Thomas: I cannot get enough of cryptography. That is for sure.

Eric: You got the bug. Exciting stuff. And so, was there something specific that ultimately got you, or can you tell us maybe your personal experience with that moment that it just really clicked for you with cryptography?

Thomas: Yeah, this colleague of mine, Ties, he told me about a multi-party computation and what really got to me was basically the counter-intuitive functionalities that he was talking about. So, it is, to me, it was at that time, very strange that you, it would not be possible to perform analysis or computations or data without actually seeing this day.

So how would it even be possible? And then he tried to explain this to me, and then tell me that there are techniques, cryptographic techniques that that allows to realize this functionality, and I just wanted to know more about it. So, he basically sucks me into this topic and a couple of years later, I am still enthusiastic about all of this.

Eric: Great. Excellent. So now TNO has a relationship with the Dutch government. Tell us a bit about that and techruption.

Thomas: Techruption. Yeah. Okay. So, the relationship with the Dutch government. First, I should actually mention or emphasize that TNO's an independent research institute. So, we have a little bit of a strange position in the Netherlands and that is negative.

We are founded by law. So basically, in the Netherlands, there is a law that, TNO law, that says there should be an institute, such as TNO, to perform an applied research, but it is also very important that we are independent from the government because we sometimes also get asked to validate or assess certain decisions or projects from ministries and those kinds of things.

And then for them, it is really important that we have an independent role. So that is, that is our relation to the Dutch government. And one of the main goals of TNO is to create innovation ecosystems in which government industry and academia collaborate. So, we really want to stimulate innovation and basically

our role as an applied research institute is to bridge the gap between academia and industry and academia and government and bring more fundamental knowledge to practice.

And well techruption corruption is an example of section innovation ecosystem. So that is like a collaboration. I think there are like 50 different parties involved in this collaboration, and within this collaboration, there are a lot of different projects in which we try to work on, on different technologies.

And techruption specifically focuses on a number of topics, which are a blockchain technology multi-party computation, but also artificial intelligence, and then self-sovereign identity, for example. So, this is just an example of one of the ecosystems in which TNO plays a big role.

Eric: Excellent. And so now, maybe we can talk a little bit about those projects and actually, maybe even before that talk about, about TNO, how do we move like multi-party computation, which is a focus area, from something that is theoretical or even purely technical into practice.

And maybe you could elaborate on that a bit.

Thomas: Yeah, of course. So, well the, the first multi-party computation protocols already stemmed from the ITs from the mid-80, and they are quite old already, but back then the protocols were definitely theoretical and not very efficient.

So, they were not applicable to many real-world situations. So, what we see the last 10, maybe even 20 years, we, what we see is that these technologies have been improved significantly. So, a lot of research efforts, academic research effort has gone into improve in these technologies, the efficiency thereof, and then thereby making them applicable.

And I think a big example is from 2009 where in Denmark, a multi-party computation was applied to facilitate sugar beets options. So, there is like a famous example of it showing that these technologies are becoming more and more applicable, and since then, since 2009, the applicability has improved much better, or so we see this development going on.

Eric: Do you, do you think those improvements were related to the scalability, the processing throughput? Do you think it was the interfaces, the usability? Do you think it was the underlying technology was enhanced? Where do you think the greatest improvements were made that really facilitated moving from something that was purely theoretical to something that was starting to become practical?

Thomas: Yeah, so at first the main improvements were scalability. Improving the performance of these protocols. So, when you apply multi-party computation protocol, you always have some sort of overhead. So, it is always a bit harder to do the computations in a privacy preserving manner than if you would not want to do these protocols without caring about privacy.

And this overhead has reduced significantly. So, if you start really technical and enhancements have taken place, in making these things applicable, and that is something that is still going on. So, people are still working on these techniques and making them more efficient more applicable.

But what we are also seeing right now is that if we want to apply these techniques and we do not only encounter technical challenges, but also legal challenges, ethical questions to ride governance issues. So, these are all sorts of non-technical questions that you encounter when you want to try and apply multi-party computation in the real world.

Eric: And so, do you want to maybe develop on some of the legal, ethical and governance considerations?

Thomas: Yeah. Yeah. Sure. For example, so let us, let us take a look at the legal considerations that is that play a role. So, what MPC basically allows you to do, it allows parties to analyze data in a privacy friendly manner.

But it still means that MPC or the application is processing sensitive information, and a lot of privacy regulations are in place to sort of prevent the analysis of sensitive information or user sensitive information, so you need basically, you need to feather around for being able to being allowed to process sensitive information.

So, in Europe, for example, we have the GDPR, which is a privacy regulation and in the Netherlands. We have other sorts of regulations that sort of prevent you from just processing sensitive information. So, there are different ways to get a valid ground for processing sensitive information. One way it could be consent.

So, if the data subjects, they give consent for you to use their data, then you are allowed to do it and you can, can perform your analysis. So that is, that is, that is one way to go about it. Another example is that some institutions, they have a legal task to perform certain analysis.

And, and this tasks also give them this legal task also gives them a valid ground for processing sensitive information. An important example is in the Netherlands, Dutch banks are, the banks are obligated to prevent and detect financial crime, for example, money laundering and fraud and those kinds of things.

So criminal money flows that the Dutch bank to try their best to detect these criminal money flows. And if they do not do this appropriately, then they can receive huge fines. So, these fines range, and can go up to hundreds of millions of euros. And the last couple of years at least two Dutch banks, two big Dutch banks have received these kinds of fines.

So, they have this legal task of detecting financial crime and then requires them to analyze transaction data. So sensitive information that the transactions between their customers, for example. So, this gives them valid ground for processing information, but even if they are allowed to process into the, this information, there, there are still restrictions what they can do.

So, for example, the Dutch banks cannot just share their information with other banks. So, they, they have to keep this data in house which already limits them in what they can do with this data. Because they, they cannot see the transactions that do not take place within their network, for example.

So that already limits their view on, transaction network. So, these are legal considerations that you really have to take into minds when you want to apply multi-party computation or when you want to

process sensitive information, whether you do it with, or without a cryptographic or privacy-enhancing technology.

Eric: So, for example when you, we typically hear about these privacy-enhancing technologies use cases being compelling in healthcare research and, and certainly in on the financial side for regulation, those are probably some of the most compelling, although there are other compelling uses for analyzing data statistically. And so, in one context you still have to get consent, but using this data can be a minefield. If you, if you inadvertently disclose it, if you want to share it with others to greater facilitate your ability to analyze the data, you run all sorts of risks. That could be preventative from a business, from a research perspective, and then as you shift to even where there is a, at the outset, a permissive use of that data in the context of complying with regulation even then you can trip on it and inadvertently disclose it and still be subject to huge fines, as well. So, you have this permissive and somewhat permissive, but even when you have right to conduct analysis, for anti-money laundering or otherwise, you still run the risk of, still having a breach and still being liable.

So, you know, particularly as we get more complex and more data-driven, it seems like a better use case is to figure out, hey, how do we just insulate us ourselves from this potential risk versus, trying to manage it in a way that may just, long-term not really be manageable. Yeah. So interesting.

Thomas: And then what multi-party computation can really do or, or other privacy enhancing technologies for that matter is realize a much higher level of proportionality. So, in a traditional approach where you do not use these privacy-enhancing technologies, data shared copies from one place to another place and then distribute it in a lot of ways.

If you if you analyze this data in a privacy preserving manner, you are processing the data in a much more proportional manner because you are not giving or grounding access to the data via by other parties. So, then that is an important step or a value that is the most part, the population, or other techniques as to this.

Eric: And I guess that sort of bleeds into one of your other points. In, in managing the legal risk, or you have to manage the, you have to manage governance, right?

Yeah, definitely. So, one of the things that for example were multi-party computation or federated learning and all these sorts of techniques, the main focus is to protect the input data.

So, the sensitive input data and to perform a sort of analysis without revealing this input data, but the outcome of the analysis can also be sensitive. So very simple, so this is something you have to do be very careful about. So, for example, if I compute the average salary of three persons and the answer of this computation or this analysis is 1000 Euros a month.

So, the average salary of three persons is 1000 Euros a month. And I already learned that no, and then no individual earns more than 3000 Euros so I can perform the analysis in a privacy preserving manner, but also the result of the analysis reveals something about the individuals. So that is something you have to be careful with.

And another example is if a fourth person joins this computation, and we recompute the average salary now not of three but four persons. And then it is very easy to deduce the fourth person's salary by comparing the two different averages. So, these are typical things that are not automatically taken care of when you apply multi-party computation.

So, when you apply multi-party computation, you also have to think about what kind of sensitive information can be juiced from the outcome of the computation. So that's a, that can be challenging to quantify this for example.

So great. And so obviously it is not just the technology, there is a lot of other considerations, as well.

In fact, we did an episode with Rand Hindi, and I said, what are the big areas to think about when we talk about things like, fully homomorphic encryption or multi-party computation? And he said, like once we have hit the holy grail and his point was, I do not know what comes after it.

He goes, but the policy considerations and how we grapple with, how do we manage? Who has what, how they use it, what can be deduced from it? All these policy considerations as to the use of the technology, the legal, the ethical, the governance, those will remain and, to some extent the law and everything else will need to catch up with the technology and everything and all the challenges that it presents.

Yeah, I guess that is where us lawyers come back into the picture.

Thomas: No, you can play a big role in that that part.

Eric: Yes. Yeah. So, shifting gears to the readiness of the market today for these technologies and I think maybe multi-party computation seems to be the more market ready, but what is your, well, I certainly do.

There is a range of market ready, but certainly I think where you see a lot of talk about practical applications, seems to center on multi-party computation and some use of multi-party computations with other privacy-enhancing technologies. But do you want to maybe elaborate on that and what from your side?

Thomas: Yeah. So, I think these designs are actually quite exciting. So, for example, in so the last couple of years we have done various projects and applied multi-party computation to two different domains. But all or most of these projects were like proof of concepts demonstrators demonstrating the applicability of multi-party computation trying to spread the word.

Look at these techniques, they, they are becoming a market ready. You should think about applying them. But what we are seeing nowadays is that these proofs of concepts are moving to real-world applications. So, for example, recently we deployed our first, ~~at Dino, we deployed the~~ first multi-party computation system on real patient data.

So, we have to there is almost this project or there is this project where a collaborates within an insurance company, a hospital, and a statistics company. And they, they, they there, they are question

was how we can analyze or determine the effectiveness of certain treatments, certain medical treatments, medical procedures.

And of course, they run into all sorts of privacy issues because if they want to assess or do this analysis, they have to analyze patient records and also information about patients and it is distributed over your different parties. So, some information is kept at the insurance company, other parts honorary information is kept at the hospital.

So, this, this, what is a typical scenario were multi-party computation applicable. And the solution that we developed was a pilot that earlier this year on a real patient data, and the pilot was actually quite successful. So, this will be further developed, and actually in a spin-off company that was recently founded Link Site.

So that is actually quite exciting to be able to tell you this today that, that this is happening, and simultaneously we also see that in 2020 another a startup was founded in the Netherlands Roseman Labs, also a company that that is aiming to apply multi-party computation to all sorts of real-world problems.

Eric: So, it is quite exciting to see all these developments and see that we are moving from proofs of concepts to actually to real world applications.

Yeah, congratulations, by the way. You know, this was, I think announced just a few weeks ago in June. We are actually, we are still in June, so you know, very recently, certainly when we are talking about in this podcast, we will try to bring it back to that practical experience.

Cause obviously you were involved in it, and it was not theoretical, it was tested and then you guys actually implemented it in in real, in the real world in a real context and, exciting. So, we talked a little bit about any money laundering or fraud detection as use cases.

You know, there are other use cases, as well. Some of which are, you talked a little bit about treatments and, and I think, for certain treatments that may be even more compelling, maybe do you want to expand a little bit on that? What you have seen, what kind of research you have done,

Thomas: Yeah, sure.

So

that HIV can appear in, in, in millions or maybe even billions different kinds of mutations.

So, there are a lot of different varieties of this specific Asia that this virus that that are around and all of these different mutations, all of different varieties can react differently to treatments. So, and, and all the other end, we have a lot of different treatments.

So, I think there are currently over 70 different treatment regimens available. And one U station can, can act, can act very, or react very positively on the certain treatments. And these treatments can be very effective for one week station with, can be very ineffective for another mutation.

So, prescribing the right treatment is a delicate procedure and a very difficult task. So, what happens nowadays is that we have these systems, clinical decision support systems, the systems are in place and doctors can use these systems to prescribe the right treatment for that patient.

So, they basically fill in all the information they know about their patients. What kind of mutation is that then? The system can say these treatments will be effective for these specific patients, but these clinical decision support systems are not. So, they, that they are there's room for I would say significant improvements.

And these, these systems can in particularly be improved if they would have access to more data, more information in particular about the successfulness of prior treatments. So, if a system prescribed certain treatments, then it would benefit from learning from the successfulness of this treatment. Then it can, for example, the next time either prescribed the same treatment or another one.

But this is not, this is currently not possible, because this would require the system to analyze patient data, which is sensitive. So, because of privacy reasons, and because we want to protect the sensitive patient records, it is currently not possible to incorporate this additional information in these clinical decision support systems.

So, what we basically did is we developed a proof-of-concept implementation, where we used information on prior treatments to improve the decisions by these automatic systems and thereby assist doctors in prescribing the right treatments.

And so, something like that obviously the more doctors participate, whether in the Netherlands or otherwise you know, outside the Netherlands, obviously more as you grow network the better the data becomes. Yeah.

Eric: It is not something today too much in hospitals, correct. Like most it is like conferring with colleagues, but all these different little pockets of research you know, start to become all that much more valuable when you start to tie it in across countries. Right?

Yeah, definitely.

And do you, where do you think the first mover will be? I know that's a, it's a difficult question, but where do you think the first mover in terms of what treatment is going to be so compelling or is going to have enough you know, international support and presumably private companies, maybe even research organizations, who knows, maybe research organizations will lead it, but where do you see any movement sort of this coalescence or are we still in an area where we're still testing it out, we're still researching it.

We are still prototyping it, and that expansive view of trying to get, trying to actually get all this data in, obviously there's scalability issues, as well. But do you see any sort of international movement that you think might actually gain some steam, sort of that, that first use case, internationally, you know,

Thomas: Yeah, so I fully agree. So, if we would be able to collaborate internationally, that would be beneficial for these kinds of systems, because more data is - gives you more insights. So, on the one end,

we see that in some of the projects we are still piloting and prototyping these, these kinds of technologies and we are demonstrating within the Netherlands, for example. So, within our borders. And the reason is that we have also we first have to make sure that people also trust these technologies, that they, they understand that if they apply these technologies, that their sensitive data is protected. So, we are also still in the face of spreading the word getting them, to become more familiar with these technologies.

And because if we would, if we already do that internationally we would get a lot of additional challenges because we have different, legislative frameworks in the different countries. And then that, then that would be, would introduce all sorts of other challenges.

So, on the one end we are doing that within our borders, and I think other countries are, we see the similar kinds of prototype proof of concept and implementation. But on the other end, especially in the academic worlds, we do see a lot of international collaboration.

So, for example, if these European projects where we collaborate with parties in other European countries and another good example, which I was not personally involved is for example, a sister, the personal health trend. So, the personal health train that is sort a system that is, it is not focused on multi-party computation, but more on, on federated learning.

And that is a project in which our hospital and the university were involved with also hospitals in Italy. So, there we do see this international collaboration already starting to be formed and we used, but then the, there when you cross borders, you introduce additional challenges.

So that at some point we have to overcome it.

Eric: do you foresee like the federated learning model being one that facilitates the, you know, consumption and analysis of the various, these various projects that use multi-party computation to develop this analysis and then share them through a federated learning environment.

Thomas: Yeah, that is, that is, that is one way we could see things develop, but I know for sure that there are some examples where it is exactly the other way around, but we basically have this federated learning architecture.

So federated learning is another privacy-enhancing technology. That that does not necessarily rely on cryptographic techniques, but the idea, the main idea of federated learning is that you bring the analysis to the data, not the data to the analysis. So instead of collecting all data that has, that is distributed over many places to, in one central location and then training some sort of model forming machine learning, or other kinds of techniques that especially if you did traditional approach, what we now do is we keep the data separated and distribute it and we train the model locally.

And what, on one of these data sets. And then we continue with the model to the SEC's second dataset, and we update the model based on this new data. So that is another way sort of preventing sensitive data from being shared between different parts. That is, different from the cryptographic techniques, multi-party computation and encryption that we talked about earlier.

And for some applications, federated learning we can be much more efficient because it does not involve any complicated cryptography can be can sometimes be quite costly. The downside of federated learning, however, is that the dots reveal or, or some information is shared between the different parties or the different data sets.

So, the data itself stays local and it is not distributed amongst the parties, but the model that is strained based on this data is shared between the different parties. So, the model parameters go from one party to the other and the other, the second part, and learn something about the underlying date of the first party based on these model parameters.

so that's sort of, a tradeoff between the different techniques. And what we are also some experiments where instead of sharing the model parameters directly with the other parties in the plane, so that they can see these model parameters, we can also share the model parameters using MPC or using homomorphic encryption.

So, then we shared a model. The model is trained locally. It is being encrypted, for example, with homomorphic encryption scheme, then it is moved to the second part. And second part, you can update these model parameters without having to see the model parameters. So now we see already a combination of federated learning and multi-party computation and that is that that is that that seems to be a very a strong combination.

Eric: Excellent. No, thank you. So, we talked about HIV, we talked about some of these other treatments, any other ones like maybe, that you think are particularly compelling for projects that are currently implementing these privacy-enhancing technologies?

Thomas: Well, there's production an interesting one that I did not think about before, but so we, we were also working at TNO on this project combating poverty, basically.

So, what we see in the Netherlands is that we have this additional benefits or health benefit or sorry, social benefits for elderly people that live below a certain minimum wealth standard. So, if they live below what minimum wealth standards today are entitled to additional, social benefits. However, these elderlies are often not aware of the social benefits.

Therefore, they do not apply to them, and they do not receive them, and as a result they are unnecessarily living below these minimum wealth standards. So, when we want to combat poverty, we want to reach out to these people and tell them, well, you are entitled to the social benefits. So come and get them basically, that is what we want to do.

But the problem is government agencies do not know where the government agencies that are responsible for the social benefits. Do not know which people are entitled to them because to know that they have to they have to have access to the income information of the Dutch citizens, and it would be very disproportional to give all the income information of all the citizens this specific government institution.

So, then they can just filter and find out who are entitled to the social benefits. So, this would be very disproportional and that is not allowed to do that. So, what we are working now, we are working on a

multi-party computation solution for this problem where they can find out who are entitled to these social benefits and without having to access the underlying data, the income information of the all the Dutch citizens.

So, this is, I think this is a very interesting application of multi-party computation that we only found out because we were talking to the to the government agency, and then they explained this problem to us, and we thought we know a solution for this specific problem.

Eric: Excellent. And also, there, Link Site has something called the Care for Data platform. Do you want to describe what that is?

Thomas: The, the spin off company recently founded that we just talked about, their main application is to analyze the effectiveness of health care of certain medical interventions and to do, to stay, have developed the Care for Data.

Well, that is a multi-party computation platform that allows them to perform this analysis in the privacy preserving manner. And their platform actually makes use of a homomorphic encryption scheme to perform the analysis in a privacy preserving manner. But it is not only the homomorphic encryption scheme that is incorporated in their platform because their platform also they also developed a governance structure.

So, the analysis that is they day when they want to be able to do our statistical numbers, and the statistical analysis that they, if you implement them nicely, then they are easily misused. So, for example, one of the parties can ask a lot of different questions on the data. So, it can, I want to compute the average for all men in this data sets.

And then I went to compute another average. And if you, if you make a lot of these intersections, then what you can do is you can do sensitive information. So, you can basically use all the information about one specific individual that you were targeting. So, this kind of misuse should be presented.

So, what they did is they, they also built a governance structure on top of their multi-party platform. So that is all part of this Care for Data, and that prevents misuse of the system. And what they actually do is they use a public ledger technology, so blockchain technology, to record the number of queries and the actual queries that the party makes to the system.

And it also limits the number of queries, the number of questions a party can ask to the system, to the data set. So, there, we see a really nice application. Next, we have multi-party computation of blockchain technology. That is make sure that for the multi-party computation protects the privacy and the blockchain technology makes sure that the system is not being misused.

So, we have accountability, basically all the actions for accountable by right by even by third parties outside of the city.

Eric: And so, where does the, the homomorphic encryption come in? How does it so the multi-party computation is the collection of the data, you know, I do not know if it uses the federated learning, as well.

But then, and then the blockchain also ensures that you do not have abuses. So, it is recording. It provides for auditability, which is a really important feature of any of these schemes. But, but how does the homomorphic encryption factor?

Thomas: Okay, so typically a multi-party computation, and when we speak about multi-party computation, where we are referring to a specific type of protocols based on secret sharing.

that achieve dysfunctionality that allow you to analyze data in the privacy preserving manner. Homomorphic encryption has a slightly different functionality. You can encrypt data and then still perform computations on this data. But there are also techniques that allow you to use homomorphic encryption to implement a multi-party computation protocol.

So that is actually what is going on under the hood of this specific solution. So, we want to have a multi-party computation because we have three different parties that want to perform this analysis collaboratively. But the, the, the way they do it is by using homomorphic encryption. So, it is basically a technique to implement this MPC functionality,

right?

Eric: It is a way of managing all the data streams you get from the multi-party computation in a way that is continues to be, encrypted in privacy preserving.

Okay. Great. So, layered. And, and when you look at the different toolkits that you have obviously you have homomorphic encryption, you have got multi-party computation. You have got a federated, you have got trusted encrypted environments, which has it have their own issues. Do you even have things like, zero knowledge proofs, how do you figure, you know, which, which we have not actually covered on this podcast.

So, if you want to actually try to tackle that one, you can, but if you are not using it, then there is no point we will save it for another episode because it gets- zero knowledge proofs is a whole world into itself. I totally, and we definitely need one on the show. So how do, what are the primary considerations that you look at?

Like when you have your bucket lists when I am making, when I am faced with a project and I am thinking about it structurally, what are the factors that drive you to say, okay, this is, these are the tools that we have to use and even how do we layer them?

Thomas: So that is an interesting question because it fits really well with what an applied research institute is doing. So, we try to match these two. So, we have all these techniques, all these tools developed by and in academia. And then on the other end, we have practical, real-world questions and challenges.

And how do we match these? How do we find the right tool for the job issue as you mentioned? So basically, what we do is we always go into a project where there was a blank sheet. So, we do not decide beforehand which technique we are going to use. But we do, we have techniques in our in our toolkits multi-party computation, federated learning.

And then what we do is we, we discuss with our partners. What is the problem that you are facing? What kind of analysis do you want to? So, typically the, of course the question is we want to analyze data, but this is sensitive information. So, so, so there are all sorts of considerations that come into play then the, the main one is what kind of security guarantees do you want?

What way? Should the privacy be protected? So, when we talk for example, about homomorphic encryption, we have these homomorphic encryption schemes that are well relatively efficient, but they are not secure against quantum computer. So that is something we should take into consideration when we want security.

So, for example, if we want to deploy a scheme that is going to be used for a long time, or that is going to analyze data that you. Secure for a long time, then we should already start thinking about the threats of quantum computing. And then we should not deploy a scheme that we know will be insecure once a quantum computer will rise.

So that is something that has to has to has to be considered. When we also see is that some of these protocols, so that there, there is not really a silver bullet, so there is not one protocol that is best for all solutions. So, some protocols require quite significant or induce quite a significant computational output.

So, they require a lot of computational powers, strong computer, strong surface to perform the analysis, for example. So, then it can be a downside, but on the other end, you have other protocols that can be computationally relatively efficient. So, they do not encourage this huge computational output, but they do require interaction between the parties.

For example, if we see homomorphic encryption, the nice thing about fully homomorphic encryption is that we can create a sort of an MPC functionality without having to communicate a lot. We just have to encrypt the data, send it to one party. This party can perform the analysis on the encrypted data.

It can take a while, but it does not have to communicate and then send the answer back to the or the result back to the input parties. So that is a nice thing about it's very convenient in some practical applications. But in other applications we cannot have this computational output we, we, we have to compute sort of analysis.

We have to perform an analysis instantaneously. Now you can, for example, look at secret sharing based multi-party computation protocols where the computations are much less involved. They can be performed much better. But the protocols do need interaction. So, we have multiple parties involved in this protocol, and while they are performing the computation, they have to keep interacting.

So, you can only perform a multi-party computation protocol when all these parties are aligned and if they are distributed over the world. So, one party is in the Netherlands and the other one is in New Zealand. For example, then we also have quite a lot of latency between the two the two parties, and that can make certain MPC solutions in applicable.

So that is also something we have to take into account where you want to apply different technologies. So, the main message that I guess is that there is not one silver bullet. We still have to really performance on some cryptographic engineering to, to find the right solution or the right tool for them.

Eric: Right. And, and the point that you made about post-quantum computing and the risk of a fully homomorphic encryption scheme, it, you know, if it is built for a long term, it is really interesting. So that suggests that where you are thinking about privacy impact of, like you are going to the government and saying, hey, we have got this great solution and privacy.

And it is Ooh, I am so excited. And then post-quantum computing, and it is like this big honeypot, to, to be attacked. Like you, you place a lot of trust into the homomorphic environment because you believe, based on today's tech you are fairly confident that it is the best solution, but then lo and behold, it becomes the very thing that is a target later on.

And so, it sounds like in that case you are thinking more in terms of you know, maybe federated learning and multi-party computation where you are just pulling the results, but the data is residing in all the sources. So, you do not have this single honeypot of risk. Is that one way of framing it or I am sure it is, I have made it too simplistic.

and that is, that is the right way of framing it.

So basically, we have all these sorts of threats that you have to take into a garden to mitigate this stretch. You can deploy different techniques and combined them, and not only to mitigate threats but sometimes it is also much more efficient to combine federated learning with multi-party computation then to just deploy the, the 90th protocol in a multi-party computation architecture directly, for example.

Thomas: So, this is also something we do and yeah.

Eric: Excellent. Excellent. No great stuff. So, one of the areas since you have this sort of practical experience with it. Do you maybe want to share like one experience you had where you came into a project with one set of assumptions, you had the discussion and you were like, okay, this is what applies, and then that thing that came out of left field where you were just like, okay, wow? Like I wish you would have thought about that sooner. Like that key learning moment. Do you want to you know, obviously every organization, learning from, from challenged assumptions over time is really the strength?

Nobody gets a, I am giving you it out. If you share something, it is part of the learning process. But anyway,

yeah. So yeah, indeed we have quite a lot of these learning processes that we find out. That is, we added a role in the beginning and then you have to visualize a little

bit.

That would be great if we always got it right the first time, right? Yeah. I guess it is not the second time.

Thomas: No, I can give an interesting example. So, a while back, we also did this project, and we were working on optimizing the workflow in our hospital. So basically, we had a hospital and they wanted to optimize their work.

So, to make sure that patients do not have to wait too long and then optimize the entire process within this hospital and what they wanted to do to achieve this is they wanted to use location data and this, so basically all the staff members. So, all the doctors, nurses, and everyone working in the hospital, they received a batch with sort of a GPS stack in there so that they could track the location of their employees during the day.

Same thing happens for patients visiting the hospital and using this data. They wanted to match the movements from patients and staff members and optimize the workflow within the within the hospital. But you can imagine that for example, this location data is very sensitive, especially in your location data of the staff members, because this location data shows whether a doctor is assisting or a specific thing, a patient, and then reworking, or whether it is standing at the coffee machine and then taking a short break.

So, you can imagine that the staff members do not want to share this information with the hospital their employer. So, what they did is basically they make sure that this data is managed by the union. So are the union they trust, and the union can see cause location, data of the of, of the staff members of the hospital.

But to optimize these workflows what a what they had to do is they have to combine this location data of the staff members with the location data of the patients. And this patient data was held by or was cabinets by the hospital. So, we have two different parties, two different data sets, and to perform the analysis, we have to combine the data sets, but for privacy reasons, we cannot just do that in a traditional manner.

So of course, MPC allows the privacy front of the solution. So, it made a lot of sense to start thinking this, like working about a multi-party computation prototype for solving precisely this issue. But when we started doing that, we, we sort took a closer look at the application and the computation that actually has to be performed.

And we found out that we could redesign the algorithm and perform certain computations locally. And by doing that, we could perform the entire analysis without having to share any data without having to use MPC at all. This was really a result of taking a closer look at the exec algorithm and then that sort of dissecting it and we are arranging the algorithm and finding out that multiplied the computation was not required at all, that you can just perform the computations locally.

So, there was actually a quite interesting, we started off with a project and the goal was to apply multi-party computation to solve a privacy problem. But in the end, we managed to solve all of it without multi-party computation, actually a really simple manner. So that is also something that, that is good to take into account.

So-so dive into the algorithm and then see whether you can do something more efficient than just nicely applying a cryptographic program.

Eric: Interesting. It, points through maybe some of the issues also with fully, with homomorphic encryption, which is you have all these different levels of centralization, right?

Where are the computations being performed? How much of the data is being retained, all that stuff? And it seems like in this case, even though multi-party computation is not more privacy preserving and not as centralized. And again, it is my own word for centralization as maybe a homomorphic and encrypted scheme, what your solution basically was about decentralizing or localizing that data more and not aggregating in a way that, that raised some of those concerns.

So that is a very key consideration, right? So, it is, it is.

Thomas: a lot of these techniques such as fully homomorphic encryption or multi-party computation, maybe it is federated learning, it can be applied in a black box manner to an algorithm. So, we have wanted to compute an algorithm or do an analysis on sensitive data.

Okay. Then we plug in MPC, federated learning or fully homomorphic. But when you do that is typically not going to be optimal. So that is going to be incurred quiet, quite large overheads, and they are often much more efficient ways of performing this analysis. So typically, you need to redesign the algorithm a little bit to make it more suitable for real multiparty computation or any of the other techniques.

Eric: Excellent. And so, another application, I think TNO has done some work on it, is the Netherlands Comprehensive Cancer Organization. Do you want to maybe touch on that and talk about some of their, the unique challenges and where it differs from the HIV and some of the other research that you have done?

Thomas: Yeah. Sure. So, the difference here is maybe the Netherlands Comprehensive Cancer Organization. Well, they are very large research institute when a lot of data locally available, so they already do all these analyses and um, that they already have a lot of information at the hands, but there are some data sources that they cannot access yet.

So, even this organization that has already the oldest information and has already before this analysis, and then also they are so a lot of data just give consent for this organization to analyze their data because of the cost is to improve, the treatments of cancer. So, a lot of people give consent for this organization to, to process their information for this this.

But there are some other factors that influence a patient's chances for survival for example when, when they have cancer, and they do not always have access to this information. So, for example, the drug usage of patients. So, whether they have an history of drug abuse or other kinds of are their other kinds of abuses or having those three conditions?

So, some of these, these are examples of information that could play an important role in the effectiveness of the treatment of cancer for which they do not affect us, too. So, when we are trying to

do with this organization is also to develop framework a system that allows them to perform this analysis and to, to also incorporate data from other organizations in there.

So, what, where we do not really have one specific computational analysis that we want to do. This should be like a multi-party computation framework that, that should allow them to do a generic analysis, on our own cancer patients.

Eric: Interesting. So, it is you know, also actually today I am doing another podcast with Somos and Genobank and they basically collecting DNA, genomic information on indigenous peoples and forming a database that would allow people to identify, certain tendencies are and health risks.

But sometimes the challenges of doing this like in a multi-party computation environment are high. And so, in their case, they are not employing homomorphic encryption, their view is that they will manage the database itself you know and protect it. And so do you think that I guess my question is a multi-party computation environment, something that is easily overlayed were maybe.

The desire is just simply to get something to market, to pull it in from so many disparate resource resources first, and then maybe figure out the privacy later. Like, what are your, what are your thoughts on that? Certainly, in Europe where I think there is a greater understanding of, know, there is you know, more of an emphasis on privacy.

So, these considerations factor in right out the gate and it avoids some of the issues that could develop later on. But what are your thoughts about, maybe in jurisdictions or countries or, where, where that may not be quite as important and in the interest of bringing something to market that has a public good.

You know, maybe they do not deploy these techniques.

So, I first I, I fully agree with you. So, what we have seen the last couple of years that privacy has become a topic in many public debates. We have the GDPR. So that is the renewed two regulations. They were renewed in 2016, I think. And in 2018, they, they became effective in the, in Europe.

So, these, these prescribed how to well, we will how handle sensitive information and data. So, this is quite an important topic in in the Netherlands and in, in an early stage already in an early stage, when I am working on new technologies, new um, new platforms where data is being analyzed, we are already forced to think about privacy considerations.

So as an example, we, we typically, work with privacy by design, to already in the design phase, we have to think about the privacy challenges and, and investigate what kind of mitigating measures are there. So, I think that is also the way you should go about it because it was always very hard once the system is up and running to modify it and to plug in crypt over fee or honor as a, at a later stage.

So, I think that is that that is going to be much more challenging than when you think about this in response, because sometimes you do need a structural change in how you go about analyzing data. So, for example, you might not want a central location to store data anymore. You might want to have a distributed data storage where, where every bar, where data is cut into different parts, such that none

of the individual parts are intelligible, but only when you bring them together, you can perform your analysis.

Thomas: So, so I think that should be the way to go about. But still sometimes there is a tradeoff between the functionality that you want to achieve and the costs of these privacy preserving, technologies. And I can imagine that there are some applications where the costs of these technology are still a little bit too high to, to, to, to, to apply them directly.

Eric: Yeah, for sure. You could even see maybe a world where, once you have post-quantum computing capabilities, like the ability to crack into some of these database in ways that we thought were previously secure, it would be interesting to see is you know, for example, we often talked about insurance providers and the ability of insurance providers to discriminate and deny coverage to people who need it based on even some sort of genetic propensity and You know, which, which would have, certainly it would be very hard on people who are particularly, poor underprivileged or don't have access to it to then get it because then they have this added layer.

And I guess you could see a world where, you have a bad actor collecting this information, but then somehow cleansing it in, in, in selling it to you know, something that appears more legitimate and say, hey, insurance provider, there is this database you could provide that, we collect the whole source of information from X, Y, Z.

And who is going to really actually go and pull the threads on all of that and say, hey, we provide this whole database to you. And it helps you identify, where, you you're, you're likely to suffer a loss because you insured somebody who had a predisposition you know, So it's to your point it's, it's something that, I would anticipate at some point in the future, there'll be some sort of event that draws that into sharper focus and then companies that are, particularly, in Europe where there's a greater emphasis are going to be better positioned to execute on, on, sort of that growing awareness and say, okay, now we get it.

This is not the result we wanted. How did this happen? And, and then, and then it sort of drives it forward. I think Europe gets it maybe America, a little less. So, you know, Americans a little more likely to sacrifice privacy for, for, for better outcomes or most people do not even, they are sort more immune, but privacy by design is definitely more of a European focused, initiative in the U S I guess growing, but certainly not where Europe is yet.

Maybe in California. I do not know.

There is always a balance. There is always a bit. Yeah,

Thomas: I agree with you, but you gave a nice example. So, for example, in this privacy regulations, that is also something stated as the right to be forgotten. So, and then, and this is very hard to realize the traditional manner where data is centrally, stored and shared with other parties.

And if you want to be forgotten, if you want to have your data record removed, for example, for some reason. So, for example, if someone has made a mistake, so if someone has said, well, I think Thomas is a criminal, for example, and it is, it puts this information one day directed in this data record to share it.

With another, institution and this institution shares it with three other institutions. Then it is a very hard for me to rectify this because I have to track down all these places where my data is stored and where it is indicated that I am a criminal and I will have to convince all of them that is not true that there, that there was a mistake made in the beginning of the face of the chain.

So, then it's very hard to realize in this, the traditional winner if you apply MPC and I and other privacy announcing technology data is not shared amongst different parties. So, it is, it is much easier to keep control of, of information.

Eric: For sure. Well, listen, this was a great discussion.

Thanks so much for coming on the podcast. If people want to find out more about what TNO is doing, more about you, and even Link Site, I do not know, to what extent you are still engaged with that, but certainly TNO is where, where can they find it?

Thomas: So, I am indeed. So, I am, I am not really involved in the Link Site but a Link Site as a website.

So, you can find more information on their websites. And also, at our TNO website, we have a lot of information about different projects that we have worked on. There is also contact information available. So, feel free to reach out if you have further questions, should be defined.

Eric: And what about you? What about people want to follow some of the work that your -things that you are working on or commenting on.

Thomas: So, so when I am working on, so most of the work that I do, I have to try to publish as, as much as I can about the work that we are doing.

So, it is always good to receive feedback and to have other people is also learn from the things that we have learned within our projects. So also, at the TNO websites, I would recommend looking around over there. You will find me, as well.

Eric: Excellent.

Thomas: Easily.

Eric: Great. Thanks so much for coming on.

It was wonderful to have you as a guest.

Thomas: Yeah. Thanks a lot Eric. I enjoyed our discussion.