

TRANSCRIPTS

2026 Videos 2026

*A complete transcript of every recorded talk from the 2026
Videos.*

Contents

Full transcripts of all 2 recorded sessions, in the order they appear in the archive.

01	2026-05-05- Introduction-to- KERI-Part-1- XYohZTh3yEw	<i>2026-05-05-Introduction-to-KERI-Part-1-XYohZTh3yEw</i>	48:48
02	Its-KERI-or- Surveillance	<i>Its-KERI-or-Surveillance</i>	54:25

TALK I

2

2026-05-05-Introduction-to- KERI-Part-1-XYohZTh3yEw

2026 VIDEOS · TALKS · 48:48

Welcome. My name is Samuel Smith. I'm the creator of the KERI suite of protocols. And this video is a tutorial talking about KERI from a high level. KERI stands for Key Event Receipt Infrastructure. It's been in development for several years now.

And you can think of it and as a secure identifier overlay as a trust spanning layer for the internet. So, we're going to talk about what that means today. For resources, you can go to the website [KERI .one](https://keri.one). Should be easy to remember. Look at the resources page. There is an active KERI community with various standards and specifications. You can find the code, which is is Apache 2 licensed at the Web of Trust GitHub repo. The Trust Over IP Foundation, which is a member of the Linux Foundation's decentralized trust organization, governs the specifications for the relevant protocols. These are licensed using the Open Web Foundation license, which is a freely licensed open specification license. And you can find meetings there and there are a couple of links for that. The motivation for KERI came from years of working with protocols. I started life as an academic, did a lot of work with AI and autonomous vehicle systems back in the 90s. Started working with protocols in the Internet of Things also back in the 90s. And later in my career, started working on cloud-based automation systems and was confronted with the problem that the internet did not have a really good security system. And if you think about it, the internet has lots of problems today and they're getting worse.

And so this idea that the internet was broken in can be traced back to the fact that the internet never had a security layer. If you look at the comparison of the OSI model and the IP model, the layers where you do authentication are in the presentation and session layers and the IP protocol doesn't have those layers. The result is that instead we use bolt-on security systems or sometimes called security overlays. These are commonly known as domain name systems, certificate authorities, openID connect, OAuth, various types of systems that provide security. And that security is based on what was a good idea 20 or 30 years ago when it was created, but now is no longer such a good idea. And so KERI is about reimagining what the internet would look like if we had a security layer that actually worked.

The idea can be traced back to something called a spanning layer. So, the internet in its early days wasn't viewed by everyone as something that was going to be universal. It was an experimental protocol. But very quickly, relatively speaking, it dominated all of the other communication protocols and became the dominant protocol for the whole world. And one of the analyses as to why that happened is that the internet had a spanning layer and that spanning layer made it possible for the internet protocol to grow in a way that allowed it to subsume all the other protocols and provide all the services. And that what that spanning layer in the internet is the IPv4 for, you know, IP version 4 and now we have IPv6. But originally this is the diagram that was used to demonstrate that there was something called a spanning layer. And if you think about it, it looks like an hourglass and the waist of the hourglass is the IP layer. This is the addressing layer. Everything above that layer uses the same address and everything below that provides support for that addressing layer and that feature of the spanning layer. And you can look at some of the papers that have been written about that is reputed to be the reason one of the reasons the internet became so successful as a protocol. The problem that we've had is that because the spanning layer of the internet did not have any security built into it, it didn't have a trust basis. What has happened is that we now have what I call platform lock trust.

In order to have trust, you have this IP spanning layer and then you have supports for it. And then you have applications built on top of the spanning layer, but they don't have trust in them. And then you bolt on trust above those. And so what you get is trust domain based segmentation of the upper layers.

And so you don't have trust across those those walled gardens or columns above that. And so you don't have trust that spans the internet. It just spans a given platform. So you get a bifurcated internet trust map. There's no trust spanning layer in the current internet. And you might think of this as like, you know, different platforms that have different trust bases or, you know, and often trust is the identity system, right? So you have a Facebook identity system and within Facebook you have trust and on Google you have its identity systems and within Google you have trust and you might have other platforms like blockchains that have trust within their trust domain but they're not interoperable. You can't move trust between those trust domains. So the basis of the spanning layer, there's a really well-written paper that talks about the hourglass approach to building protocols where you have a spanning layer. And so we feel that that's the right approach, but we just need to have a trust spanning layer. And we don't want to reinvent a new internet that would be ludicrous, but we want to allow the internet to remain the internet, but add a spanning layer that allows trust. So there's not locked to different platforms. And so our solution we call a waste and a neck is we have the IP spanning layer, which is the waste, and then we add a trust spanning layer that is the neck. And if all trust spanned applications use the same trust spanning layer, then they'll have interoperability in terms of their trust. And so you can still have platforms, you can still have all the protocols down below, but if you want to have applications that span trust, where trust spans across the applications, then we want to have a trust spanning layer. So that's the design goal of KERI.

So I've used the term trust basis or trust domain a lot. And so I just want to define it so that you understand what I'm talking about. A trust basis is the set of trust roots. And when we talk about systems design, we talk about things like roots of trust, things that you can rely on that have certain well-defined properties. And then based on those things, you can build other services and functions on top of that. So we talk about a trust basis and then a trust domain is the ecosystems of interactions that rely on a given trust basis. So for a trust spanning layer, we would need to have everything above that rely on the trust basis that is the spanning layer. And so KERI is about defining what that looks like. So, the basic idea of KERI, and we'll go back into this in more detail, but the idea is that we can define an identifier system security overlay. It's a new type of overlay that acts as the trust spanning layer. So, you have infrastructure that sits below the secure identifier overlay or the secure identity overlay. And then you have a trust domain that sits above that. And the spanning layer is that overlay. Now, when people look at KERI, often they look at it and it's difficult to understand. And one of the reasons that KERI is difficult to understand is that we call it the platypus of the identity space. And the reason for that is if you look at the platypus, it's a very interesting animal. It's difficult to classify because it has a unique combination of features not found elsewhere in nature. or at least not in that combination. For example, because it has milk, a platypus looks like a mammal, but it also lays eggs, so it looks like a reptile. It has a bill, so it looks like a bird. It has electrolocation, so it looks like a ray. And it has venom, so it looks like a snake. And so if you approach a platypus from the mindset that every identity system has to look like a mammal, it will be confusing. Or if you approach it from any of the other viewpoints. So KERI is best approached from the fact that it is truly unique in many important ways.

So KERI deconstructs what many people are familiar with in identity and security systems to create a novel combination of properties that solve what we call the universal secure attribution problem. And we'll go into that in more detail. If you think about a system design trade space, you can have lots of puzzle pieces that you can put together and there's lots that one can select from. But KERI carefully selected

a subset and left certain things out that it didn't need so that it got the properties that it did need to solve the problem and the problem of a trust spanning layer requires solving the problem of universal secure attribution.

And if you think about it, if I want a trust spanning layer, then I have to be able to do something universally. So KERI is a well-fit-together puzzle where the pieces are carefully designed to match that trade space. So let's step back a little bit because we've talked about the internet. Let's talk about each individual person. Why should I care about having a trust spanning layer on the internet. Why should it matter to me? The fundamental idea that KERI tries to solve is to enable entities, both natural persons and legal entities of all types, to retake control of our data and the value of that data. And if you think about the internet and how it's evolved, most of the value of the data on the internet is not controlled by the individuals that source the data or subjects of the data, but they're controlled by large internet companies, super aggregators, super data correlators.

They control the value of the internet. People gain a lot of value of interactions, but many of those interactions are designed to maximize engagement and often engagement is maximized through rage or other deplorable human type interactions. And so we want to have high value interactions on the internet that are valuable to the people interacting. They're not being exploited.

So when you think about building a system, you think about, well, how do I build something? What are my tools? What are the toolkits available to me? And so if my toolkit, and these are tools from my tool chest in the garage, these are all hammers, various types of hammers. and they work really well with nails. If I have a set of screws, those hammers aren't very useful. I need a different toolkit. I need screwdrivers. So we need to figure out what our toolkit is because actually in security systems, we don't have a choice of toolkits. We've got one. So what we have to do is we have to figure out how to make it work with that one toolkit and we have to be able to build our systems with just one So if our toolkit is screwdrivers, we've got to build everything with screws. If our toolkit is hammers, we have to build everything with nails. So there's really only one set of tools for truly secure data control. We only have one. And so we have to figure out how to best use that one. And I would say that the existing internet security systems don't do a really good job of the existing toolkit. Now, fair enough to say that 20 years ago, ago, the toolkit was even worse than it is now. They had fewer tools and they did the best they could with the set of tools they had then. But we have better tools now, but it's still of the same type of tools and we could do better than they did. And so KERI is about recognizing that the tools have gotten enough better that we can do significantly better in our toolkit. And those tools are entropy-derived tools. So these are primarily cryptographic one-way functions.

So these are things like cryptographic strength pseudorandom number generators, hashes or digests, the ECC key pair generation algorithms, digital signatures, encryption. Now, in a few years, we're going to have to add another set of tools because quantum computers are going to provide the ability to tax ECC key pair generation. And so we need to have post -quantum key pair generation algorithms, which NIST has now approved. So it has to be on a roadmap. But for today, we can just focus and KERI was designed five years ago. So we can focus on that, but KERI is cryptographically agile. So KERI doesn't have already has a migration path to be post-quantum secure. So let's talk a little bit about the economics of data. As I said, we want to we want to be able to retake control of our data and its value. So when you think about economics, it's about value, right? And it's about the creation and capture of value. And creation and capture looks like extraction, exchange, and exploitation, both in bad and good ways. And so if

we want to take control over the value of our data, data, then we need some way to secure that control. So the goal is control over the value. And there are two types of value. There are its, these are atoms. People are built out of atoms. So you can think of people as value of people as also atoms. We have things built out of atoms, buildings, cars, things. And to secure those atoms, we have something called physical security. And there are lots of physical security systems and tools for physical security. And for the most part, it's pretty difficult to attack physical security. It usually involves putting lives at risk to ultimately gain physical control over atoms that are protected.

Physical security is largely a solved problem. The other type of value are bits. This is information value. And the problem with information is that it's digital and can be copied. So you have documents, you have places to store data, but you can copy information at will indefinitely without losing any of its fidelity.

Whereas we don't yet know how to copy atoms. We don't yet know how to clone people. It requires new atoms. You can build things that look like old things, but they're different things. And the cost of building them is high. Whereas information, to make a digital copy of information, the cost is extremely low.

And so information security is by nature a different type of security because we can't have a single copy that we protect physically. We have multiple copies. And once those copies leak out, then they're everywhere. So, and then the question is, you know, what does that look like? You know, from the standpoint of people, you know, how do people control, you know, how do we have information security? So, we want control over value. We have the two types of security. And the one thing that we have, in our toolkit, and remember I said they're entropy-derived tools, is that we can have digital uniqueness. It's a very narrow type of digital uniqueness and it's called entropy. And basically we can create numbers that are derived randomly such that somebody can't guess the number as long as the number is kept secret. So if we don't share the number, they can't guess it. And that is the only security that we have. Everything else is one-way functions that try to preserve that original entropy, that original level of cryptographic security that says they can't guess the number. And so that's not a very rich tool set. So now how do we control things? Well, we as people, we are its. We want to control bits. How do we do that? Well, the way we do that is that we create cryptographic pseudonymous identifiers that are derived from entropy and we build a mechanism that allows us to prove that we control those identifiers. We're not talking about all the information in the world. We're talking about just identifiers. We can control identifiers. And if we can control identifiers, And then if we have a way to make attribution to information to an identifier, then we can at least assert that the source of the information came from the person or entity that controlled the identifier. That still doesn't sound very satisfying, but that is all we have.

So, we have to build from that. So, let's think about it again. to retake control of our data, we will first retake control of our identifiers. Now, this is where the internet falls down. If you have a domain name, you're renting it. You don't own it.

You're renting it from a service provider. There's a limited number of those domain names, so they're a scarce resource. And so nobody can own it. They just rent it. All right. Now, you can maintain rental control indefinitely, but it's still not yours. If you create an account on some social networking site like Facebook and you get an account there, you don't own that account. Facebook owns that account.

You can create a username and a password, but you don't own what's on that account. And that username and password, they can at any point in time say, we don't want you on our account and they can remove you from the account. And so you don't have control over the data and you don't have control over the identifier. So since the only thing that we can have information security over are group graphically derived identifiers, we must first retake control of our identifiers. So that's what KERI does. KERI says, we're going to develop a system where each individual controls their identifiers and hence their identity. When we talk about identity, we're going to narrowly think of it as identifier plus attributes and anything that can be securely attributed to that identifier. That's the identity. So it's a very narrow definition, but it's the only definition that actually works. If we define it more broadly, then we don't have control over it. And so there's lots of differences in identity. So I tend to avoid identity in general. And when I say identity, I really mean an identifier system security overlay, or I might call it an identity system security overlay, but it's really about the identifiers. So identifier uniqueness comes from captured entropy. Anybody can capture entropy of sufficient strength that there is no ability to attack it. Even a quantum computer can't take that away from you if you do it right. So what we create are cryptographic pseudonymous identifiers. Those are the types of identifiers. We call those cryptonyms. That's short for cryptographic pseudonym. And the hard problem is how do we provide provable perpetual control over that cryptonym?

That's the hard problem KERI solves. It's not ephemeral, not temporary, but perpetual control over a cryptonet. If you don't control your identifier in perpetuity, then it's not a very useful identifier. Now, I want to go from the identifier to the data. If we have the ability to have provable control over identifier, what we would like to do is we like to have control over data associated with that identifier. So the first thing we want, and I call this the four A's of secure data control or A raise to the fourth power data, right? So we need an author of the data, the creator, the source of truth, the root of trust for that data. And that imbues the author is the only one with the authority to control that data, right? So authorship and authority are if you can establish that, then you've created the basis of secure secure control. Now you need to be able to authenticate any action that you perform.

So if you create data, you need to be able to prove to somebody else that the origin of that data was from you, the author, right? So you need to be able to do secure attribution. You need to be able to provenance the data. When we talk about provenancing a painting, right? What's the provenance of a masterpiece? Well, the way you provenance it is that you trace the ownership of that painting back to the original painter, the original author of the painting, the painter. If you don't have a chain of custody of that painting all the way to the present day, then you can't securely establish that the painting is not a forgery. It may be a forgery.

Forgers are really good. You can look at the painting and you go, well, it doesn't appear to be a forgery. There's no obvious things that make it a forgery. So that's a good sign that it might be authentic. But without the provenance, the chain of custody all the way back to the painter, you don't know for sure that it's an authentic painting. So we want provable origin. We want provenance. We want secure attribution. And then if we've created data that we can prove is sourced from us, then we want to be able to control what happens to that data. And that means authorized the use of that data, give permissions, consent. It establishes what people call a loci of control for the data. That is the domain over which you have control. So if the data is in your loci of control and you maintain that loci of control, then you have control over the data. And so you need mechanisms to authorize the use of that data or what happens with the data or anything else.

And then finally, a property that is not cryptographic at all is authoritative. For data to have value, it must have a creator, it must have a provable origin, the use of it must be authorized, but the data also has to be authoritative. It has to be credible, accurate, reputable. And so we need mechanisms for that which we can bootstrap on top of the other three.

Let's now go back to the trust basis of humanity when we interact as humans and think about how we interact as humans because we're built really well for those sorts of interactions. Our brains and our senses and how we function, our culture is all has been designed to enable us to function really well with the human trust basis. And in the human trust basis, we interact with people in person. We see them, we touch them, we hear them with our physical senses. And over time, we get to know people from those in-person interactions. And so that knowledge, right, that we as a human can know other people means that we can then trust them.

And how we trust them is based on those physical interactions, right? As humans, we're interacting. And so we build the ability to trust other people and to trust what they say and what they do, right? And it doesn't mean that we aren't fooled or tricked or people don't lie. All of those things happen. But we have a mechanism by which we operate fairly well that works pretty well, has worked pretty well throughout history. And we have lots of institutions that bulwark that human trust basis. The problem is, is on the internet, we can't really know somebody human use in the human trust basis because we don't interact with them physically. We don't see them with our senses. We don't touch them. So we can't really trust what comes over the internet because for all we know, the bits that came over the internet didn't come from a real person or didn't come from the person that said they came with. They could have come from somebody else or they could have come from an AI or for a bot or from a dog for all that matters, right? It is a much more difficult problem and we don't have the tooling as humans to do that. We have to use different tools. So we have to be able to switch from a human trust basis to the only tool that we have for information security, which is a cryptographic trust basis. For those of you who don't like cryptography, this is probably not the thing you really wanted to hear. But a cryptographic trust base is the only one that we have. And to the extent that we can use a cryptographic trust basis, then we can have really strong information security. To the extent that we use other trust bases, then we're not going to have strong information security. So one of the goals of KERI is to provide strong information security without requiring everybody to know anything about cryptography. You just need to know the best practice for using it and hopefully all the tooling under the hood hides all the cryptography from the person. You know, this tutorial is to lead up to for people to get to understand KERI at increasing levels of depth. So this is really high level theoretical abstract, but we're going to start going deeper as we get further along. So one of the problems we have to recognize is that cryptography has limitations. For example, one of the cryptographic tools that we use heavily is a verifiable digital signature. And it depends on something called a cryptographic public-private key pair. It's also called asymmetric crypto. So with a digital signature, you can take a statement. This is a digital statement.

You sign it with a private key and you can verify it with a public key. This makes it very useful for communicating over the internet. And it enables the strongest form of security because you're not sharing a secret. And what does that mean?

mean? Well, with a digital signature, any signed statement is non-reputable by the controller who signed it. What we mean by controller is the individual entity or natural person that controls the private key, right? Because in order to sign it, you have to have the private key. So, what does non-reputable

mean? It means the signer cannot plausibly deny that they created the verifiable signature.

Therefore, we may only verify the controller of the private key and we call this the who that made the statement. So the author, right? But not the veracity of the statement. We can't use cryptography to prove the veracity or accuracy or truth of a statement. And that's the what, right? But we can prove the who to at least the level of plausible deniability. Okay. In order for that to occur, the assumption is that the private keys behind a digital signature, the private key or keys are truly private, never shared, right? And that the signing, the act of signing reflects the signer's true intent. And we call this these two things together, the signing infrastructure. And we say the signing infrastructure was not compromised. So, for example, if somebody steals your private key, then they can sign stuff and impersonate you. They can say they are who you are and somebody trying verifying it will say, well, it verifies, therefore it must have been signed by the private key. And so if the private key has been stolen, then you lose non-repudiation because now anybody, whoever, multiple people now have that private key. Correspondingly, if I have a computer or a device that uses the private key to create the digital signature, I have to have some way of inputting to that device something to be signed. And so that digital artifact that I input to be signed, if somebody is able to input something to be signed and I permit that signature to happen, maybe it's on my device and they get it onto my device and they tricked me into signing it, then now I've signed something that I didn't intend to sign. And so that's another type of compromise. So the assumption is that either the keys or the signing or the act of signing has not been compromised. If that is true, then digital signatures provide non-repudiation to the signer. And that gives you a cryptographic basis for securely attributing a statement that has an attached signature to the controller of the private key as a source of that statement. So with cryptography alone, we cannot trust what was said. It doesn't provide that. We have to use it to other mechanisms in concert with cryptography, but we can say something about who said it. We can only detect if the signer is being inconsistent or duplicitous. So what that says is if you don't know what was the statement is true or false, but you do know that the signer made the statement.

They're basically saying, I say this. So someone could say the sky is green or they could say the sky is blue. Well, if they're consistent and they always say the sky is green, they may be wrong, they may be misinformed, they may be delusional, but they're not being inconsistent, they're not being duplicitous. If on the other hand, they say the sky is blue one day and the sky is green another day, they're not consistent. Or if they say the sky is green to a set of people and the sky is blue to a German set of people, they're being duplicitous. And if they sign both those statements and you produce a statement that says the sky is green and the statement that says the sky is blue and you see both statements, you verify both of them, now you have provable duplicity or provable inconsistency depending on how you think about it. And therefore, you have reason not to trust the signer. But if you never see inconsistency or duplicity in all of the signer sign statements, then you may be able to trust that signer. At least you know that they're not being inconsistent. They're not being duplicitous. They still may be delusional or they still may be malicious and be a liar, but they're a really good liar because they always use the same lie. And that's something. That's not nothing. That's really, really valuable on the internet because then we can use other mechanisms to hold them accountable for the veracity of their statements. And those other mechanisms now are mechanisms that are already established. We have all sorts of legal, regulatory, and economic mechanisms and incentives that we can apply to incentivize people to have veracity in what they say. And those incentives are enforceable if we can prove who said it and that a given person or entity said something. And so we start there and then we can finish with stuff that isn't cryptographic, but we can't get there over the internet without the cryptography because over the internet, it's just data.

It's not its. So what makes KERI possible? KERI was not possible 20 years ago. It was possible probably only 10 years ago. And actually the development of KERI, I started working in this space heavily in 2015. It was possible a few years before that, but not likely. And KERI was a result of five years of trying to solve the secure attribution problem and failing. and then learning from all my mistakes. And then in 2019, I wrote the first KERI white paper and 2020, the second edition of the KERI white paper, which is largely the KERI we see today.

And since then, we've been building on that. So what happened in, you know, between 20, before 2015 and after 2015? Well, then before 2015, it was generally considered by everyone that users could not manage private keys. What I mean by cryptographic keys, to manage them, to store them, create them, not lose them, use them to sign things. That was too hard. In fact, the DNS certificate authority system and the federated identity, OIDC, OAuth, all of those systems are based on the idea that the average user can't manage private keys. So we get federated identity largely with things called IDPs or identity providers.

Identity providers are enterprises or institutions or entities that have the capability, they have the education, they have the technical know-how to manage private keys. And everybody else is a relying party because they're not managing private keys. So we built a system based on this assumption, the impossibility of user-managed private keys. It doesn't work very well anymore. Now, post-2015, we have mobile devices. We have mobile phones that are essentially supercomputers in your pocket that have devices and, you know, hardware security modules, multi-factor authentication, secure boots, secure enclaves, password management apps. And a lot of that was actually a result of all of the excitement over blockchain that started in the big inflection point, I would say, is 2015. That all of a sudden now, because all of blockchains and cryptocurrencies are based on users managing private keys. And so users were forced to figure out how to manage them and people were forced to develop tooling to manage private keys. And so that generated a whole bunch of effort that resulted in private key management. So when it came to solving the identity problem and creating a trust spending layer, I could say, well, I can assume that users can manage private keys. So I can assume that people can control identifiers that are cryptographically derived from private keys and therefore they can use those private keys to do secure provenance of those identifiers. So now I have the ability to do that. So 2015 practical user managed private keys. And since 2015 15, it's gotten even more practical. I mean, the user experiences has gotten much, much better. People now regularly use passkeys, which under the hood are just public-private key pairs.

This is something that we shouldn't say anymore, that we can't build identity systems where the basis of the identity system is on an outdated assumption, the assumption that users can't manage private keys. That's outdated. We can build them in And we have devices that do that. So then we had to settle for good enough security from federated identity. And federated identity has lots of security problems. Primarily because it's based on shared secrets. And shared secrets are reputable. And now those shared secrets are phishable or spirit phishable or any number of attacks. And so, you know, a shared secret is a password. A shared secret is a OAuth bearer token. Shared secret is a TLS exchanged cryptographic key that is used to do a TLS handshake to provide what is called authcrypt. It's a weak authentication mechanism that's reputable and we shouldn't use anymore.

But the whole internet is based on it. So that meant that with good enough security, you have to have secure sessions. You ended up with perimeter security with VPNs that also use shared secrets because you Every user of the VPN is using the same private key because the IT admin is sending out the same

client certificate to all of the users. Now we can have identity based on user managed private keys. And that means we can have a difference in security that is different in class because we don't have to ever use shared secrets ever again. And KERI is that. KERI is a protocol designed to maximize that. So let's talk about private key management. What that means is that you have controlling keys. You have key pairs for a digital signature. The controlling keys are the private keys. You create an identifier. We haven't said how we create the identifier. We'll get into that. But abstractly, you create an identifier from those controlling keys. You now have data and you use that identifier and some cryptography, namely digital signatures, to establish secure authorship or attribution over that data. The problem is that the continued use of private keys expose them to what are called side channel attacks. What that means is that they're not actually attacking side channel means they're not directly attacking the digital signature. The digital signing algorithm, if you're using elliptic cryptography, especially the most secure ones like ED25519, It has been secure for the last 20 years. And multiple Moore's laws increases in computing power have not materially weakened the security of those digital signatures. So elliptic curve digital signatures are secure. Now, when quantum computers come around, that won't be true anymore, but we already have new cryptographic algorithms that replace ECC that give us the same thing as just the key sizes are bigger and take more computation. But it's still not vulnerable to direct attack. A side channel attack means you're not breaking the crypto, you're breaking the key management infrastructure and you're breaking the signing infrastructure. And so the more you use private keys, the more you use your signing infrastructure, the more likely an attacker is to figure out how to break your use of it, not the cryptographic algorithm, but the way you use it and the systems you use it on. And so what has happened is what happens then is that practically speaking, the keys become weak from exposure. The more you expose them, the higher the likelihood that one of these side channel attacks will be successful. So the way to manage that, so this is the hard problem with key management, is that your private keys need to be revoked and replaced with new private keys. The short term we use that for that is to rotate your keys. You rotate from an old set of keys to a new set of keys. And the new set of keys give you a fresh start that haven't been exposed to CIPHERHILL attack. And so now the attacker has to start over, right? He has to do whatever to try to compromise your system. And so if you rotate them often enough, you can stay ahead of the attacker in general. Not always, but that's the general practical thing. The problem is most systems that are based on public key, if PKI stands for public key infrastructure for the PKI system is based on public keys, you must reestablish the root of trust with each rotation because the root of trust is based on the public key, not an identifier that persists, but the public key. So you throw away the public key when you rotate it. Now you have to create a new public key and you have to redo all of the work you have to do to reestablish the root of trust of that public key as being authoritative for a given person or entity. That means the identifier is how you link that person to whatever it is that you want to security attribute. And if you throw the identifier away, which is the public key, then you have to reestablish all of the reputation. All of that has to be reestablished. When you have a compromise, you break the chain of trust of control over the identifier.

So if the private key is compromised, then the public key, the provenance is gone, everything's gone. So you go, well, wow, that's really hard, right? I have to do private key management if I want, if I, you know, to be able to do cryptography.

But now the hard problem is not the cryptography. The hard problem is the management of your private keys, right? So key compromised recovery is how you fix this problem. You need a mechanism that despite your key management breaking down, can you regain control? Can you recover from that? So the

hard problem of private key management is key rotation with key compromise recovery, right? So that even if your private keys are compromised or your signing infrastructure is compromised, that you can recover control over an identifier that persists despite key rotation. KERI solves this hard problem. Until KERI came along, nobody solved this problem. The other people have now copied KERI and some parts of KERI.

But in terms of a full system that solves this hard problem, as far as I know, KERI is still the only protocol on the planet that is public and open source and open standard. There may be proprietary ones that I don't know about, but of all the public ones, it's the only one that I know about. That's the end of this discussion.

So hopefully I've whet your appetite for understanding what KERI does because I basically told you how KERI solves the hard problem with key management. And if we don't solve the hard problem with key management, then we can't control our identifiers on the internet, which means we can't control our data and we can't control value of our data. So if we really want to control the value of our data on the internet, we need to control our identifiers and we need to do key management to do that and to do that we need KERI. So I hope that's sufficient motivation. Thank you and remember to go back up to the first slide has links where you can go do some more research and look for additional tutorials on this topic.

TALK II



Its-KERI-or-Surveillance

2026 VIDEOS · TALKS · 54:25

With the KERI Foundation. I'm also the originator of the KERI suite of protocols which are gaining adoption in various sectors. I wanted to have a discussion today about what I'm calling the last best hope for individual control digital ID in an AI world. And it was prompted by the recent release by the Five Eyes surveillance conglomerate between those countries about the cyber security threat that foundational genAI models pose and contrasting that with what KERI provides and KERI stands for Key Event Receipt Infrastructure.

It's a technical term but basically it's a digital identity system that retains individual control over their identity which has some important properties. So I want to talk about that in more detail. This is an excerpt from the report, the link is there, from the NSA. It's saying that frontier AI models are anticipated to exceed current industry expectations.

This is a very technically-worded warning basically in months not years these AI models will overwhelm any cyber security capabilities of existing systems and so they're basically telling everyone in this article (it's a short article) you need to change your stance. You need to, change how you do things because we can't guarantee that your current security systems will survive what's happening. And what is happening is that certain foundation models, Mythos for example is one were developed to try to provide better cyber security, but in the process of training them to be better at detecting exploits, they also became good at exploiting exploits. And so that means that other less, shall we say, virtuous parties are building comparable models. And once those are available to cyber attackers, then pretty much the current way that we do digital identity, which is what digital security is based on, will be broken everywhere. And that's a problem. And so we need to start doing something about it. What that means is, is that we actually have two paths forward.

The two paths are 1: AI based digital identity and 2: KERI based digital identity. When I say KERI I mean the KERI suite of protocols. There's a bunch of protocols bunch of different things that are associated with KERI.

When I say AI based digital identity, I'm talking about what the Five Eyes Report recommended is that people start using AI-based counter measures, defensive measures. In other words, it's like a <i>spy versus spy</i> sort of thing where both sides have AIs and the AIs are working really hard to either defend against cyber attack or to engage in cyber attacks. And so they're saying everybody needs to start using AI to defend themselves because the AI attackers will be able to defeat your current identity security systems and that's a potential problem. Now KERI, as we'll talk about a little bit, was designed to be highly resilient in this sort of scenario. So what does an AI-based digital identity look like? Well, it's a digital persona that is derived from continuous tracking and statistic correlation of one's behavior and biometrics expression. Basically, the AI creates a digital picture of your behavior based on tracking and correlating your behavior and then an attacker has to impersonate something that correlates better to your behavior than your own behavior. And so as long as your AI is doing better job of correlating your behavior and collecting those signals called fraud signals or anti-fraud signals then the AI can say no it's you, and not an impersonator, whereas the attacker is saying no I can create an AI that can impersonate you, can deep fake you, can make it look like it's you, and so what happens is that everything that you're protecting digital identity AI learns how to do to better identify you, the attacker will eventually learn how to better impersonate you. And so it's just a continuous cycle. And the only way that you win

is if your protection AI has as much information as it possibly can have about your behavior so that it can do a better job of creating a digital persona that matches you. So if that isn't scary from a privacy tracking surveillance perspective, you don't know what scary is because that is pretty scary.

But that's what that looks like. Whereas KERI is based on a cryptographically-derived fault-tolerant survivable ID (red.) and it's survivable to AI cyber attack. We call it perpetual ID. It isn't based on the weak things that current internet-based identity are based on which are largely shared secrets. It's not based on the weaknesses that are in the DNS certificate authority system and TLS and OAuth and OIDC and all of these systems that we have developed over the last 30 years that are now vulnerable to AI cyber attacks.

It's a ground up new design that is designed from the beginning to be fault tolerant to survive attack to be able to detect compromise and recover from it. So it is designed to be resilient to these these sorts of attacks.

So what happens is that if you go with AI-based identity you can get protection but inevitably it's going to be totally centralized with zero privacy. Whereas if you go the KERI route, it's fully decentralized with best possible privacy. And that doesn't mean that you have the privacy that maybe you thought you could have. It means it has the best privacy you possibly can have given we're living in this new AI environment where AIs are consuming all of the data on the internet and there have available to them, data to correlate pretty much anything about anything you do.

So we can talk about it in more detail about what best possible privacy looks like. So what we want to do is in this discussion is answer some high-level questions about what this means. So these questions have topics and then there's a question and then I'm going to go into more detail on that. But let's go back and remember we're talking about AI-based digital identity which NSA and the major security organizations from the Five Eyes countries have said (implicitly, red) "everyone needs to start adopting" or the alternative is a KERI based digital identity. KERI's not as well known, so people don't know about it.

Although it's got some significant adoption vectors, including one is called State Endorsed Digital Identity (SEDI). The other one's called Open Verifiable Calling (OVC), which is getting rid of spam in calling, and in the healthcare sector, there's work on using KERI for patient ID. All of those are major significant adoption vectors that that could put KERI in a position to be the last best hope for individual identity because if we go down the AI route, we may get security. It's an arms race. It'll always be an arms race.

But we won't have any semblance of user control or privacy. So, let's talk about that. The Five Eyes statement calls for AI assisted defense. So the question is how do we ensure AI remains subordinate to humans, law, policy, and cryptographically verifiable identity infrastructure rather than becoming a de facto authority.

Well the way that we ensure that is that we need to have cryptographically verifiable identity in the first place. If we don't have that, then we can never ensure that it doesn't become a de facto authority because we can't tell what it's doing. It's not transparent. It's not observable. It's basically an AI and that AI is going to make decisions about who you are. And the goal is to protect you from fraud, but AIs themselves don't have any inherent structure that allows them to separate out what is sensitive information and sensitive authority, and sensitive things that it should do or not do versus anything that it's within its feature space or weight space, based on the the prompts that it's received and the objectives that it's

received and the temperature by which it's deciding which path to follow and I can go into more detail but OWASP recently released in March a 100-page cyber security report on the new AI threat and it is stark and because the AI threat poses a threat that hasn't ever existed before and there aren't any solutions for it. They say that in the report. All you can do is to try to mitigate it to some degree.

So, how we ensure it? First is: we have to have cryptographic verifiable identity infrastructure. So that that is the authority and individually controlled. So what are the hardest legal technical governance mechanisms that are needed to ensure that every significant AI action remains attributable or accountable to humans and organizations? Well, that means that we have to build a regulatory and a legal bull-work for our individual rights.

The original Constitution of the United States has a couple of amendments and the fourth and fifth amendments, but those were written in an age when there wasn't a digital realm when pretty much everything was on paper. Information was disseminated on paper. We now have a digital realm where information is disseminated electronically. And the way you protect yourself in the digital realm and the way you protect your identity, your information, your presence is different enough that we can't rely solely on those two protections. We need to have additional protections for that. And one of those protections is called a "duty of data loyalty". It's probably the best legal framework that's been developed over the last few years.

There's some legal scholars that have been working on it, written several papers on it and the state of Utah recently adopted it in the legislation. That requires that people who process data about you must use that data in a way that is in your best interest. And that's the start. And then you have to layer on top of that a mechanism for people to be in control of their identity to be able to then express their preferences. And it's not simply consent. It's actual engaging in what is in their best interest and attaching to the use of that data what is called chain-link confidentiality where the where the terms and the follow the data as it goes through the system so that you have an ability to enforce it so that it also includes citizen enforcement or citizen recourse.

So there's several there's layers of mechanisms that we need to put into place and we're starting to do that in some places. The state of Utah is being a leader in it and there's a there is a coalition of other states that are looking at creating legislation the same way that it gives us an opportunity to be able to do that. But absent that it's going to be very difficult to stem the tide of what's happening with AI and because of the urgency it makes it all the more problematic.

So the NSA's reports calls for stronger identity and access controls. Right. Well there's a problem there. the problem is what counts as stronger, right? I mean, the current controls and people have been using them for years and the and people keep strengthening them and they keep adding to them, but the AI can attack them in such a broad way now that they need to be strengthened in a way not ever done before.

It's like I guess you could say it's like going from armor made out of leather to armor made out of steel, right? Like to go from weak to strong, but you don't have the ability to build steel armor yet. You haven't invented steel. So, what do you do?

So KERI was designed from the ground up to not require an AI to defend against an AI. It was designed to use cryptography in the best possible way that you can use cryptography which requires key management and it requires key management done in a special way. And if you do key management in

that way, then you can build a system that is resilient to these sorts of attacks because it doesn't have the weaknesses which are built into the current controls because of the state of technology 30 years ago when these systems were designed and then they've been bolted on and added on and modified but the core architecture hasn't changed in 30 years whereas KERI is a new architecture.

so this question is which weaknesses in current identity systems become most dangerous when attackers can operate at AI speed and scale. Well, most vulnerabilities typically employ a combination of vulnerabilities. So you might have an algorithmic an infrastructure, a software and an operating system and a usage vulnerability that together enable an attack. A recursive privilege escalation attack often takes advantages of multiple vulnerabilities to be able to both move vertically and horizontally in an or in an access control organizational space by exploiting different different degrees of vulnerabilities.

And so it doesn't get to where it wants to go all at once. It just gradually works there because it finds what's broken at each point and then searches for the next exploit, right? And with AI, they can search faster. They can combine exploits faster because they can search the whole space and they go "Oh, wait..."

In fact, there was a recent paper, I think last week, where they were able to exploit the memory security model in Apple's M5 architecture through a combination of human and machine AI analysis that the humans could've never done on their own because they just didn't have the ability to analyze all of the possible different things they could think about. they didn't necessarily know how to how to finish the exploit but with the human guidance in combination so AI speed and scale is not just the AI itself it's fact the attackers are human, the most of these attacks are guided by humans, either criminals or nation states, that want to come after critical or valuable infrastructure and the AI just is a force multiplier for for all of those types of attacks, not just, but is for sure a force multiplier. And that means that even hard to find unknown exploits that are latent, that are in your system are going to be exploited quickly, which means that your system has to be designed differently. It has to be designed to be fault tolerant to be able to survive even if it's exploited. And so that is a different approach to security.

So let's talk about that a little bit more. Cyber resilience after breach because that's the perfect segway. The statement assumes breaches are inevitable. and the question is in a KERI-based architecture what remains trustworthy after servers databases cloud providers certificate authorities are compromised?

Well, in a KERI-based architecture, what remains trustworthy is that you have built a a fault-tolerant mechanism that enables you to recover to detect and recover from compromise. And so what happens is that you you may not be able to anticipate all of the ways that an attacker can attack you. But if you have mechanisms that allow you to detect compromise and then a way to recover from compromise in a timely fashion, then the attacker doesn't have time to exploit that compromise. And like in a *Recursive Privilege Escalation Attack*, usually it's a sequence of exploits that take time. Now the time doesn't have to be very long. It might be hours or days, now prior to AI, well the current the average time for a breach to get detected currently and I guess that's no less what you know and having an AI as a protector might reduce that time because it's better able to detect breaches, right? That's that's the idea. is 6 months and then 90 days after that to contain the breach. So it's nine months of time that an attacker has to to exploit the vulnerability that they've discovered.

So KERI is designed so that the detection time is on the order of seconds. It's network propagation time. So that breaches that are exploitable must exhibit detectability. So the whole system's designed around detectability, right?

Everything exhibits as *duplicity*. It's a duplicity evidence system. So for an attacker to benefit, they have to do something that exhibits as duplicity, which then enables the controller of the identity to recover from that in a timely fashion. Which doesn't mean that there aren't successful attacks.

It means that the blast radius of the attack is minimized. In many cases the ROI to the attack or the return on investment is negative. I mean it costs more to execute the attack than they get from it because the attack is quickly contained and recovered from.

So these systems that we currently use, servers, database, cloud providers, certificate authorities, they're all based on an antiquated, outdated notion of security and key management. And those are really harsh terms. And I know a lot of people in the industry don't like it when I say things like that, but they are. It's 30 years old. And you can walk back the choices and they were good choices 30 years ago or even 20 years ago or even 10 years ago. Not so much 10 years ago, but 20 years ago for sure. But at least in the last 10 years we've had better technology for key management and KERI is based on understanding that if you have better technology for key management then you do different things. All of these systems that that use DNS, TLS, OAuth, OIDC-based systems all use shared secrets and shared secrets are inherently weak and inherently contribute to *recursive privilege escalation attacks*.

And so it's really just figuring out different ways to to exploit those shared secrets. KERI doesn't use shared secrets. It doesn't have any. And so the protection mechanisms are fundamentally different, right? If you don't share something, then somebody can't capture it and collect it, right?

It requires a different type of attack. And even if they're successful in that attack, if you have a recovery mechanism, then you can recover from it. And so you don't spend your time trying to build a tank that's impervious to any sort of bomb. You spend your time building a squadron of tanks so that some of them can be destroyed, but the squadron survives and repairs itself and is able to continue to operate. And that's what *fault* *tolerant* means. So I use that analogy. You know, you can build a tank with armor thick enough that withstand any shell, but then it stops being a tank which is mobile, right? By definition, tank must be mobile.

So it goes from a tank to a bunker, right? You have to make some trades. And and the right trade in an AI world is a fault-tolerant system, not a a system that you have to throw away and start over with every few days or months because you can't tell whether or not you've been attacked or breached because you have no mechanism for detecting compromise.

So actually this question, I sort of jumped again on it, but how does KERI suite shift security from protecting infrastructure protecting secrets to reveal verifiable truths and provenance. So the idea in cryptography, in *asymmetric* cryptography, is that you have private keys and public keys; and private keys are truly private in that they're never shared.

The problem is that people who want to have a digital identity have to manage those private keys. And so that was always viewed as too difficult a problem to manage. And so we needed to have identity providers that manage private keys on our behalf, right? And then the key management was hidden from the system. So you have identity providers and relying parties and then you use usernames and passwords or now *Pass Keys* There's stronger things than usernames and passwords, but you're still

relying on an identity provider for the ultimate security of the system. And that ends up looking like shared secrets, *<i>bearer</i>* *<i>tokens</i>*, that the identity provider distributes around to people to entities to be able to engage in things over secure channels, right? The secure channels are protected with shared secrets. TLS, you know, uses a Diffie-Helman key exchange shared secret to exchange an encryption key which is then weakly authenticated by encrypting something and then the other side decrypts it and then sends it back to you. But it's subject to key-compromise *<i>impersonation attacks</i>* and so anytime you start doing things to avoid the hard problems of key management, you end up weakening security. KERI says "No, we're going to build key management the right way."

"We're going to make it fault tolerant." "And then we're going to enable people to use it in a way that requires too much effort for the attack to be successful." And KERI uses *<i>threshold</i>* structures *<i>multi-</i>* *<i>signature</i>*, *<i>witness</i>* pools, *<i>watcher</i>* pools and delegated AIDs that are hierarchical, so that you can build as secure a system as you want; so that you can tune the security to the value that an attacker has to go for. So if it's going after something really valuable, then you layer on additional threshold structures. And relative to the cost of the resources to run a defensive AI to protect your system, the infrastructure for KERI is trivially costly; in comparison. So, that's really the thing that we're looking at.

Let's talk about privacy without surveillance. Can strong security, accountability, and resilience be achieved without continuous monitoring and surveillance? Right. Well, it can. Because what KERI does, It says we're going to use cryptography the way it was meant to be used.

Without explaining the KERI protocol, which would take too much time, it doesn't have the weaknesses that our current shared- secret based systems do, which are vulnerable to exploit by AIs. AIs can't brute force a public key to get to a private key.

Now if I have a quantumcomputer that's a different story but we now have post-quantum public-private key generation algorithms. And if your system's cryptographically agile you switch to them which KERI already is. So that's a different problem.

so what they have to do is what's called a side-channel attack. They have to find a way to exploit through side channels. Whereas a shared secret multiple parties have a copy of that secret. So the types of side- channel attacks that work are much greater. If I've got a KERI private key that's truly private, it's never shared. It never has to be shared. And even if that private key is compromised, I can recover from it through KERI's *<i>pre-rotation</i>* mechanism, then an attacker doesn't gain much by compromising my private keys. They have to compromise much more and that much more can be designed to be low friction for the user but high cost to the attacker in such a way that you invert the resource expenditure.

So it's too expensive to succeed in an attack against the KERI infrastructure than what you gain for it because of the the inherent distributed nature and decentralized nature of its design. and I'll go back to this word *<i>balance privacy</i>* and how does KERI balance privacy and accountability?

Security requires secure attribution. You need to know who's doing what. But you also want to be able to control what context people know about who's doing what. And KERI does that. KERI gives you the ability to have identifiers that you use in specific contexts that are not correlatable to identifiers you use in other contexts.

But in each context, you have perfectly secure attribution to all the parties in the context. And what that means is that you get rid of the main source of fraud which is impersonation attacks which deep fakes all of those things because every participant must be accountable in that context. They can't pretend to be somebody else because you can challenge them and they have to prove that link. So if it's an AI, that's an agent working on behalf of somebody, you can challenge all the way back to that person. If you're using KERI and that person is delegating authority to an agentic AI, then you can always establish the chain of authority, right? So accountability comes from chains of accountability or chains of authority or chains of provenance or chains of custody.

and KERI inherently is designed to support those sorts of structures. Whereas most security systems are flat. They don't have this hierarchy that enables you to establish these delegations that are strongly cryptographically-verifiably chained together.

privacy without surveillance. What is the minimum amount of data disclosure needed to achieve accountability without creating surveillance society? That's an interesting question. Basically, in each context, each transaction, each application, the minimum is different because for high-stakes transactions, where there's high value, usually the amount of trust required amongst the parties is higher. That means they need to be able to better predict the outcome and the behavior of the participants to that interaction.

For low-stakes interactions, the amount is much less. So you need a system that gives you enough for the highest transactions but still allows you to tune it for the lowest-stakes transactions and KERI does that. But how do you not create a surveillance society? Well, everybody in that transaction has to have balanced obligations, right? How do you need it to achieve accountability without creating surveillance? Well, if the surveillors are accountable, then the people being surveilled will have recourse against the surveillance.

And this a change in philosophy. It's been for the last 15 years that I've been working in this space, we were hoping that there were technological means where you could hide. Well, basically you could engage in transactions and hide from surveillance. And that's not true. With AI, there's no hiding.

If the data is on the internet, it's going to be correlated. Even the smallest signal is correlatable given enough coverage; and AIs have enough coverage, right? People are building AI data centers that are going to consume a significant fraction of the world's global electrical production and what do they do? They correlate, they're correlation machines and they consume all the possible data that you can give them and correlate everything so that every point of information becomes a correlatable signal that results in something that is actionable. So you're not going to stop the correlation. What you can stop is the *exploitation* of the correlation and that means putting accountability on the exploiters.

That requires a slightly different ,well, a profoundly different legal and regulatory approach. But from an identity point of view, it requires understanding that that hiding isn't how you do it. You can partition contexts so that you can keep the interaction space from leaking out too fast, but it will leak. No matter what you do, it will leak. But what you want to be able to do is make your ability to enforce the exploitation or misuse of that data stronger. And that and so contextually, Contextually-specific identifiers can be used to remove plausible deniability by the other parties to those transactions. So there's some specific things we can do.

We're left with a stark choice. If frontier AI reaches superhuman offensive cyber capability, do our choices become increasingly stark? They they do. In fact, a lot of the discussions I've been having over the last few years are often in the decentralized identity space, and they're usually discussions about, well, we want to weaken security, so we can get privacy. We're willing to make trade-offs along those lines. And I I've always said you can't weaken security. If you weaken security, if you give up security for privacy, then you get neither. And now that's become obvious. It's sort of like: now, if you want security at all, it's really hard. And if you don't have security, then you have nothing. Because if you can be impersonated at will by an AI, then you have no identity. And therefore, you have nothing that's private, right? Because I can't assume that anything that I do can be kept private if an AI can impersonate me and prove to the rest of the world that it was me that did it.

Basically, what does that mean? See, people think about privacy as well, I can do things that nobody else can do and nobody knows about. But if someone can attribute to me, behavior that I didn't do and I can't disprove it and I then suffer the consequences of it, what does that mean in terms of me having any sort of private life? It means I don't, it means my life is governed and determined by outside forces. My *locus of control* is zero and so privacy is meaningless when when your locus of control is zero.

What does it mean? It doesn't mean anything. You're you're essentially a serf or a slave or a minion of the forces that control your identity. So it starts with control. You have to have control before you can have privacy.

And so the stark choice is you need to have security before you can have privacy and you can either have security that has zero privacy which is which is an AI-based identity system because it requires total continuous monitoring and surveillance and be able to do a better job at correlating your behavior to create your digital persona than the cyber attackers are doing to impersonate your persona. Or you use KERI which is based on a cryptographically-verifiable identity that's fault tolerant, uses key management provenance and accountability. So it doesn't solve the privacy problem the same way that you might think but it does solve it in the only way possible in today's environment.

Why is there no stable middle ground? I think I just answered that. If AI- cyber-based attack mechanisms can defeat conventional security, based on shared secrets, not KERI, then there is no privacy. It's total surveillance in order for you to have any possibility of having any sort of value in your life.

Otherwise, you can't do anything on the internet. You have to go off-grid. Totally off-grid. Like off-grid in a way that would be for most people a horrible life. This is an interesting question. As AI capabilities increase, why does continuous tracking or surveillance based trust become less sustainable than cryptographically-verifiable trust? The reason it becomes less sustainable is that it's a never-ending arms race.

The information that's electronic, once it leaks out, it never goes away. And it enables correlation to be all that much better. The correlation algorithms, the tracking algorithms, target tracking. You can think of as there's a feature space and if I'm tracking something, it's moving through that feature space. Now, in the physical world, that feature space might be, latitude, longitude, altitude and time, right? And then you got velocity, acceleration, all of those things, right?

But in identity world, the feature space has to do with things like "when you logged in, to where did you log in, what did you do, what did you say, what time was it, what browser did you use, what device did you use, what protocol did you use?" All of those are features that can be used to create a digital fin-

gerprint that then can be tracked as you move through that space. And so it forms a trajectory of your behavior. And the more information you have, the more high fidelity that trajectory is. People call it a digital twin. So they're creating a digital twin. And they want that digital twin to be as close to you as possible so that somebody can't create a digital twin that is not yours.

And your AI that's working on your behalf can tell the difference. Right? Well, that means that increasingly, as AIs gain access to more and more information, as more and more of our society and more and more of what we do becomes AI enabled, then the information available to both track you and impersonate you increases. So, what happens is the pervasiveness will always go to maximal, right?

There's no other approach and maximal means high resource, I mean we're already talking about that most of the world's electricity being consumed by AIs. Well it's only going to get more as time goes on it doesn't get less it gets more, whereas cryptographically verifiable trust is based on a completely different set of constraints that have to do with what's called computational infeasibility and exposure. And if you don't expose something, so that it can't be collected, if it's a true private piece of information and you can prove something publicly without disclosing the private thing, that's what a digital signature does. There's various types of proofs and you can prove that indefinitely or perpetually. So you have perpetually verifiable proofs. Then you have the foundation for a security system that maintains the security without getting more expensive over time. In fact, it gets less expensive over time because most of the cost is really reducing user interface friction. We're in the early days of reducing user interface friction. So that becomes a perfectly scalable system relative to the alternative Keeping humans digitally free as AI becomes the dominant interface to the digital world. How must KERI be applied to AI itself so that people remain digitally free rather than becoming exploitable, surveilled, tracked, manipulated or controlled? Well, we want AI to work on our behalf. We want the best of both worlds. We want AI that does what we tell it to do, when we tell it to do it, to the extent we tell it to do it. And for that, we need to be able to have strong delegation and have enforcement mechanisms where we can publish to other AIs that says "Our AI is only supposed to do this."

These are the rules. You can verify what it's supposed to do. When it doesn't then you cannot trust my AI. And so there are people talking about reputation systems for AIs but those reputation systems have to be managed in a way that humans control and we can do that with KERI. If you have identity systems that allow secure attribution that are not based on correlation-tracking mechanism but are based on cryptographic mechanisms. Then you can build delegation mechanisms and authorization mechanisms that are fundamentally different in structure.

They break this flat weight space constraint that AI has right now that makes it impossible to secure AI because everything's in the same weight space to where you can actually partition or segment the weight space into things that are delegable and things that aren't. Should AI be trusted to protect human freedom or should AI itself be constrained by cryptographically variable trust architecture that keeps ultimate authority with people? Well, there are a lot of people who are talking about the fact that AI can't be trusted and doom and gloom for society once an AI becomes smart enough.

I think, I spent 20 years doing research in AI systems. In fact, I did that for the early part of my career before I started working on digital identity. And yeah, we can build we can build transparent AIs that only do things that we allow them to do if we architect them differently. That doesn't mean that we have to throw away generative AI.

We just have to partition the space in which it operates so that it's one thing to have a generative AI that understands natural language. It's another thing to have a generative AI that you allow to do things on your behalf; that's an agent. So we can have a generic AI that is controlled but uses you know interfaces and that means that the interface layer has to have safety jackets and when I used to do work with autonomous vehicle systems most of the code we wrote were safety jackets to keep the autonomous vehicle alive and well and safe and to keep it from doing bad things. But we sort of forgotten about that.

People are worried about that. I mean a lot of people are trying to build those but they're building it with systems that they're using the same AI for, to secure the safety jackets and you can't do that. The safety jackets has to be foundationally different in nature from a computational point of view.

And if you do that, then you have a chance. What would a world look like where every AI agent operates under cryptographically-verified authority by delegation and thus accountability? Well, it'd be the best of both worlds.

We would be able to have AI help us be a force multiplier for us without worried about the AI hurting us. There's always going to be edge cases where people will make mistakes and AIs will be able to exploit things, but it won't be this Armageddon kind of situation that we're facing right now. The race against time. If frontier AI is advancing on time scale, months rather than years, how quickly can society realistically deploy either of these features at global scale? Well, the stark reality is the only thing that you're going to be able to deploy at global scale today is AI because AI is already deployed at a global scale, right?

But that doesn't mean that you can't switch to KERI as it becomes available to you, right? You can always make the choice to say "you know what?, I've been using this digital persona AI, but I want to switch to using cryptographically-verifiable human ..., individual-controlled identity system." And so there, as I mentioned early in the talk, there are adoption vectors. People are spending resources, but nowhere near the amount of resources going into AI systems, into AI identity system. If we had a fraction of that spend going into KERI, KERI would already be available. But we're looking at rolling out KERI for the State Endorsed Digital Identity (SEDI) initiative in the next 12 months. So, we're really close and like I said, they're already using KERI in the early adoption in the telecom sector and the healthcare sector.

So the seed has been planted. We've got shoots growing out of ground but those shoots can grow exponentially fast. All they need is sunlight, water and soil and resources. So we just need resources and and the adoption can go much faster.

Once adopted, then it becomes a bulwark against future encroachment. So, we get off the arms race. We stop being in a resource arms race. We do a one-time spend to convert to KERI and then we don't have to keep spending and spending to protect ourselves.

Which approach can be deployed more quickly, more safely, more humanely at planetary scale? Well, that's a trick question. Certainly, more safely and more humanely, KERI. More quickly? Not today but five years from now, possibly, because you don't have to every five years double the number of servers that you have running in AI data centers you don't have to fill space with AI data centers and able to keep up with the arms race. So yeah, so on a 5-year time frame, if quick is 5 years, then KERI definitely is the only viable choice.

What are the biggest technical economic political and adoption barriers to deploying SEDI/KERISuite fast enough to become a viable alternative before surveillance- based systems become entrenched? Well, one is just people don't know about it.

so they don't know to that it's that actually is a viable choice. The other is that society is filled with lots of people who are unaware of this threat and so they still think that old systems are the way to go and so they're sort of having the old fashioned arguments.

It's like they're moving moving the deck chairs on the Titanic, right? The Titanic's going down. It's going to go down in an hour, right? As soon as the whole breaches and it cracks in half, it's gone, right? But you have this sense that "Oh well, it's still floating, so we've got time, right? We've got time, and there's no rush." And the fact of the matter is you don't have time, which means that you should stop worrying about anything other than these two alternatives. And if you stop worrying about all the alternatives but these two, then it becomes pretty clear where for most people that care about this stuff where their effort should go that they should be helping ..., because nothing else is good enough besides KERI. There are lots of "KERI light" or partial KERI or they do some of KERI but not all of KERI stuff out there. And the problem is that those systems aren't good enough because they leave out important things you need to protect against the AI threat. So they're half measures, right? And half measures aren't good enough, right? No more half measures.

We use that motto If you care about security, you can't do half measures. `You're either sufficiently secure or you're wiped out.` So that's the alternative. And I have this picture that I got from Chris Bramwell who's working on this SEDI for the State of Utah.

SEDI: the last best hope against AI. So it shows SEDI as a warrior, The knight to fight the AI scourge of identity. Not that AI is inherently evil. It's just that the the only solution that an AI-based identity gives us is a centralized, identity that requires maximal surveillance. It's basically you need to be continuously surveilled, so that it can build a high-fidelity persona of you that is better than the attacker's AI, who is also continuously surveilling you, trying to build a sufficiently-high fidelity that it can steal your identity and exploit you. And that's a stark choice. Neither of those are good, right? Neither of those are where we want to end up. We want to get off that train. KERI is is a path to get off that train.

I hope this has been helpful. I hope that if you're worried, if you're <i>only</i> worried, then you're not taking this seriously enough. You should be worried like you've never been before about the future of digital identity. Given how fast the frontier AI models are growing in their ability to exploit current digital identity systems which all security systems are foundationally based on digital identity systems.

You use digital identity security overlay for everything else that you're doing. If those are breakable, then we're looking at an impending catastrophe. What's going to end up happening is people are just going to adopt the easy path which is adopt AI-based digital identity.

That would be a bad choice for us. Maybe in the short run it's the only choice available to us because KERI isn't universally available. It's still a chut in the ground. Not not a whole field of plants. But it could be a whole field of plants in the next couple of years. Not in the next couple of months, not unless a bunch of resources get thrown at it, which I don't think is going to happen.

Video: Sam and Caleb Smith Subtitles: Henk van Cann and Kor Dwarshuis