Harold Byun:

Hi. Good afternoon, good morning, good evening. Welcome to today's presentation on, Why You Can't Stop Data Breaches. My name is Harold Byun, head of product management for a company called Baffle. And today we're going to be walking through some common security gaps around the threat model that a lot of people envision around securing their data. And time permitting, we'll also get into a demo towards the end, we'll see. I'd like to wrap this all up within 45 minutes. I think people are still rolling in, in terms of the general agenda and I'll just roll through the intros as people filter in so we can get going. It will be a little review of some of the mega trends that are impacting data security overall in marketplace and technology adoption, some of the common data security controls and why we see that they're not as resilient in the face of modern day hacks today, a little bit of coverage on data security and governance frameworks, we'll cover some of that and some methods that are emerging around mitigating the risks.

Harold Byun:

The demo will be time permitting and obviously QA, feel free to ask questions throughout, there's the chat window for folks who are not familiar with this. So you can feel free to chat away and ask questions. I'll try to pick those up as we go through the presentation as well. And you can always email me as well, or info@baffle.io. So hopefully this will be an interesting and engaging session for you. In terms of a little bit about my background, I've done roughly 30 years in security, both on the security architecture side, as well as the security analyst and admin side before switching over to the dark side and going on to the vendor side. But my entire career has really been focused on what I would consider a theme around data containment. So back and a long time ago in stints and data loss prevention, I worked with Skyhigh Networks as a CASB provider for a while, and have been with Baffle for three years now probably working with the company for well over four years at this point, again, with a pretty unique approach around data security.

Harold Byun:

Let's get right into this. In terms of key trends and security gaps, I don't think that any of this is new information for many of you, but the data breaches continue to happen. There's a study that was done in 2019, that stated that roughly eight billion records had been leaked from the cloud in 2019 alone. The numbers are staggering, and the breaches continue ad nauseum. And it begs the fundamental question as to why does this continue? And why can't people get better? And there's a litany of reasons around it, but we're going to look at this more from a data centric threat modeling perspective today to really look at things within the context of what I would hesitate to say our assumed breach posture positions, or and I hate to use the term but the zero trust type of scenario or posture, and really what that means in terms of access to your data. But effectively untrusted assets or unknown assets, and unknown profiles or users accessing your data, which lends itself well to speaking on number two, which is really around third party risk and data sharing. You've been in security for a while, you're very well familiar with third party risk assessments. And if you go back 15, 20 years ago, there used to be a page and a half. Now, they're several 10s, or hundreds of pages long with roughly 1500 questions in terms of the average third party risk questionnaire.

Harold Byun:

And at the end of the day, even after that and all those questions are answered and scored, it really doesn't necessarily speak to how the data is being handled, and what the operational footprint is of that

third party provider as they may be accessing your data. And roughly two thirds of CSOs have reported that they've had a leakage via third party. And then the third piece is really the cloud storage data leaks. The status thing, over 1 billion records leak. There's other studies that suggest and again, it's in excess of seven billion records in 2019 alone, and with the data footprint continuing to grow and become more distributed, we don't necessarily see that slowing down. This is just a survey that was done by the 451 group, where they surveyed a number of different security leaders in terms of considerations around data management overall and data analytics and security and obviously, data privacy. We're topping out are some of the top concerns within that survey that was taken.

Harold Byun:

The underlying privacy or data privacy regulation context continues to progress. This is just a headline around once CCPA took effect, even with the way they were handling it under COVID, I think this was probably the second week of July that this actually occurred and the actual effect date was July 1st of 2020. Obviously, the regulations are being enforced with greater rigor, and we'll make see that continuing to increase down the line as well. Beyond the meta trend of just what are the security risks and the penalties aside from just even the penalties as it relates to breaches is what we're hearing a lot from customers is also just the increased regulatory authority that comes over your organization beyond the fines. There are a lot of cases where organizations have been fined a set number of fees, I think Capital One's was $80 million. There was another LAGASCA financial services firm back in October, that got fined $60 million. And those aren't chump change numbers, but they're not really going to break the bank per se, as it relates to those organizations.

Harold Byun:

What really starts grinding things to a halt is the regulatory oversight and overhead that comes in post of those types of breaches beyond the fines and that adds an incredible amount of overhead and greatly impacts an organization's ability to continue to innovate and deploy applications and move in a more nimble fashion. And so, there are pretty some pretty strong considerations in that regard. The other trend is really around overall data distribution and cloud data lakes. There's obviously, with the launch of Snowflake or not the launch of Snowflake, but at least the IPO and the continued growth of that company, they're projecting roughly close to 100% year on year growth adoption for data footprint within data warehousing and analytics solutions, such as a Snowflake or Redshift are alternate solutions. These are just some other stats that have come from various sources in terms of how organizations are looking to leverage data going forward.

Harold Byun:

And ultimately, what we're continuing to see as well is a hard stopping point with on premise big data environments where people just cannot keep up with the data footprint that they've invested in, on premise and are looking at ways to take on a more flexible type of cloud data lake solution. And so ultimately, this translates into greater distribution and a lot of ways potentially lack of visibility and a loss of control. When you see the move towards things like a cloud data lake that often get consumed into a data warehousing or analytics footprint, you typically see a movement of something like, this is obviously just a slide where, but something that might simulate something like this, and it's a pattern that we see repeatedly where we have customers that are looking at moving off of legacy environments, mainframe environments, and migrating that into a cloud data lake. And in some cases, that's just a pool of S3

buckets that are consuming data on an ongoing basis and then pipelining that up into an analytic solution for machine learning or AI or other types of warehousing and reporting.

Harold Byun:

So this is a very common pattern that we've seen, it's no different if you're doing Azure, GCP. The former slide was AWS, this is what it looks like with Azure with people pipelining into Azure Blob storage and doing the same thing with Azure Data Lake services and Azure synapse and other analytic solutions that are available to them. It's a very common pattern, a lot of challenges in terms of controlling the identification, access control to those environments, as well as re identification of the data in terms of who can actually see it and work with it. So some of the backdrop of the overall security trends and that are potentially impacting the data security posture that your organization may have. And I'd like to delve into a look into common data security controls and why we believe that they continue to fail. We've written a multi part blog series on this, we've worked a lot in terms of secure computation and have looked at this from a lot of different angles.

Harold Byun:

And so, the first place that I would like to start and I'm not going to spend a ton of time on this slide just in the interest of time, is that there are a range or a bevy of controls available to you for whether that be on premises, these are specific to cloud providers in terms of some of the capabilities that they provide to restrict access to provide better visibility and monitoring to provide better overall control of the data that's actually being stored in the infrastructure as part of this shared responsibility model. And so, these are all things that are available to you within your collective security toolbox to help mitigate some of the risks of the these hacks. And so, obviously want to leverage these as part of a best practice and defense in depth strategy, and obviously want to look at ways that you can automate the enforcement of these controls and obviously detect config drift as well within these environments. But that's really not going to be the primary focus of this discussion, and these slides are available to you. So if this seems to be useful to you, you're welcome to download that.

Harold Byun:

What I really wanted to focus on was really what some of the focus has been on the industry in terms of data encryption and recommendations around that. And part of the reason is because when you look at compliance and privacy, there are often recommendations, but they're often not explicit. And in places where they do get explicit, as we've highlighted on the slide, it becomes pretty interesting. And in many cases, this is the IRS Publication 1075, which is based off a lot of the NIST SP publications which we'll also cover in a little bit. But really what they're talking about in many ways is, how do you secure authentication? Or how do you secure transmission or data in transit? And it's important obviously, if people are sniffing on the wire, you obviously want to have that lockdown, although SSL and TLS are pretty prevalent these days. But this is the primary mandate with nothing actually being spoken to in terms of actually how you would protect actual data values itself.

Harold Byun:

And so, if we go into the NIST recommendations and SP 800-171, and 53, or five, which is the most recent, again, we find a lot of references to how to mitigate risk and preserve confidentiality, but it's within the context of remote access sessions. So this is really looking at your VPN and what cryptographic schemes you're using to encrypt the VPN traffic. What are you doing during transmission

again, or in this case, we're in the bottom part, we're talking about wireless access points and the encryption scheme for wireless access. Everything is again on the wire. And here we even talk about within the context of data mining protection, how data is sent and received. Again, going back to transmission and everything being in transit, and I think it's important, you obviously want to lock things down every single step of the way.

Harold Byun:

And then finally, in again, another recommendation in 53, or five around transmission, so when we look at things in the context, again, going back to that backdrop of over eight billion data records were exposed in 2019, the vast majority of those did not occur over the wire. And yet the primary phone perhaps you could argue, well, it's a chicken and egg problem, it's because the wire's secured. But inevitably, you have to look at the fact that the data was hacked out of data repositories, whether that be an S3 bucket or a database that was somehow breached, either laterally by an attacker or somehow gaining access to the environment. And the point being that if the repositories of data are wide open, and it really doesn't matter how secure the wires are because you no longer need to sniff on the wire, you nearly no longer need physical access because you can ultimately get to the data source itself.

Harold Byun:

And so, that's what we see is one of the biggest gaps in terms of the overall approach for threat modeling around data security. And if you look at the latest 53 or five publication, they are now including reference to the identification of data and the ability to mask or encrypt or hash identifiable information. But this was a thing that was not present in the art for revision for publication, so it is something that is relatively new. It's a good step in the right direction, but it's going to take a while for a lot of organizations to catch up. And the reason that we talk about this a lot is because the bulk of organizations that we talk to still don't fundamentally understand the difference between encryption at rest and database encryption versus a data centric protection mechanism. And so, this is something that we still find day in day out people struggle to grasp this.

Harold Byun:

And so, for many of you who have been in the industry for many years, you probably do already get this, but we will walk through the threat model and show some different methods for how to address these types of gaps. I just realized that I didn't go through all of these. So hopefully you'll catch up again, we'll transmission and then there's a little delay here, but these are all data sent and received going back. Thank you for alluding me to this again during transmission these recommendations. And so when we go into the scenario of how most of the leaks are happening, they're happening out of data repositories. And this was that most recent de identification reference that I was talking to within the NIST publication. So when we look at Encryption At-Rest, or physical disk encryption and database encryption, the primary challenge that we have with them is that they don't really mitigate any risk in a modern day attack.

Harold Byun:

Again, going back to how attackers are getting at data, they're coming in over the wire, they're not sniffing the wire. And if they're coming in over the wire, then ultimately, they're getting into the network and moving laterally and gaining access to systems. And if they're gaining access to systems and you're using physical disk encryption, or you're using database encryption, the data is presented to any attacker

or anybody who gets access to this system in the clear. The data on that system is absolutely in the clear whether or not you're using these two encryption methods. And so, that is a fundamental gap in terms of how people are looking at this problem. We jokingly call this protection against the Tom Cruise attack, right? It's the one where Tom Cruise sneaks through the HCAC docks and drops in from the ceiling repels down, and steals the hard drive out of the most secure data center in the world.

Harold Byun:

And that's just not how the attacks are happening these days. They're coming in over the wire, a lot of times the infrastructure is living in an incredibly secure cloud infrastructure provider, or a colo. And in that regard, nobody's really stealing hard drives anymore. This was designed to mitigate against laptops that were left in the backseat of a taxicab or somebody doing a smash and grab in a storefront window where somebody left the server or under a desk or something. And so I'm not saying that those aren't valid threats, but by and large in terms of how data is being stolen today, that's not the method that people are using. And so, when we take a closer look at this, Encryption At-Rest, and I know it's a bit of an eye chart on the screen, but the fact that when you're looking at clear text data, and this is somebody who has access to the system, logging into the database and retrieving the data with physical disk encryption, and as you can see, all the data is in the clear, any of the logs are in the clear and the memory is in the clear.

Harold Byun:

And so, anybody with access to that system can grab data off of it regardless of what checkbox you selected, or what hardware physical disk vendor you're using. Same thing with database encryption, so TDE or Tablespace encryption, there's a lot of variants around this, very popular, very common, but again, if you have access to the system, the data is available to anybody with access to the system and the clear, in the memory it's in the clear, any PII or private data is written into the logs in the clear. And it really does nothing to mitigate against that modern day hack. Just a couple of data points and both Marriott and Capital One, were using TDE. And in fact, the encryption keys are present on the database using this model, which is just a commingling of the encryption key with the encrypted data container. What do we think about how people should be looking at this problem. And so, if you look at the common data access methods today, this is obviously oversimplified, but in many ways, you have a user who could be a good user, a compromised user, an unknown or a bad actor trying to get into your environment.

Harold Byun:

And they come through an application tier that has a certain set of code executing to access different data footprints in your organization. And that could be a good well profiled and known application, something where you maybe have even hashed the binaries, so that you can determine that it's a valid executable or process running in your environment. It could be an unknown application, it could be a malicious application. And this is you know how Equifax got hacked. The attacker landed onto the app tier using a struts vulnerability and placed a web shell console on that app tier. And from there, he gained access to the rest of the environment potential under the footprint. And then you have the backend access into the back end access is really a series or a collection of data that may be served up the legitimate data requests, or could be accessed by privileged users or insiders, or it could just be excessive data requests where people are trying to get additional information they shouldn't have access to or exfiltrate data in mass.

Harold Byun:

And so, when we look at this problem and how that's changed over time, when you go back to the trends that we were talking about at the beginning with the distribution of data, well, you have the third parties that we were talking about which adds to the user footprint in terms of your knowledge, you now have a prevalence of microservices and containers and how people are actually building their app jar and accessing data. There's obviously a rise in serverless code functions as well as API gateways and API access. And then you throw the database in the cloud, or you throw your data lake in the cloud. And you have a very distributed data environment with multiple access points inbound and outbound, that become a lot more harder to control. And so now we have a bigger set of users and a more distributed set of users to control against. We now have an application footprint that is spanning an entire variety of legacy and tier applications, as well as this container tier environment, or even server less code that is more transient in nature.

Harold Byun:

And then we have the cloud footprint which really opens up the door for what you would call untrusted data stores and untrusted access, and another set of privileged users that could potentially access to data, or third party subpoena from different nation states that may try and gain access to data. So there's a lot more interest in the overall international political space and what that means for coercive access to data as you operate both in the United States and internationally. And so this becomes a much more untenable problem to get your arms around from a threat modeling perspective. And if you are willing to grant or give some substance to this notion of zero trust networks, and I'm not saying I like the term, I think it's a marketing over hyped term, but if you are willing to grant that is a posture on approach to look at the world from an untrusted perspective, then ultimately, you are granting that actors are gaining access to your network, and bad actors are getting access to your systems.

Harold Byun:

And if they are actually getting that far in, then inevitably, they will reach the data. And then if your logical approach to mitigating that threat risk at that data level is to put it into a container where any attacker can see it in the clear, then you're not really taking the right risk mitigation or security control approach given the threat. And so that's the premise of how we've been thinking about this problem and the threat model and why we believe that in the current way that people are approaching, or implementing methods to mitigate those types of attacks, it's inevitable that the breaches are obviously going to continue to happen. And part of it is based on predicated on the fact that there's a fundamental misunderstanding of what those security controls do and what threats they're actually protecting against.

Harold Byun:

Let me shift gears a little bit, and then we'll hopefully talk about some other ways that we can mitigate those risks. This is something that many of you are probably very familiar with. Again, conceptually, it isn't really directly related to some of the security control gaps that we're just talking about, but it's a framework of how you might approach security or governance for your data within the modern world. And obviously, I think many people are familiar with the shared responsibility model. But just to quickly rehash it, the shared responsibility model basically states that the infrastructure provider is going to provide you with infrastructure, data center and network connectivity, and is going to be responsible for the securing of the environment. But in terms of actually implementing the controls for administrating

access to that infrastructure, as well as securing the data that goes into that infrastructure, you the customer are responsible for that portion of the security controls. And so, another view of this is a landscape of cloud security.

Harold Byun:

This is from Gartner, where they have CASB which is cloud access security brokers like the Skyhigh or Netskope, we're in that vein, CSPM, was really cloud security posture management. So folks like Evident.io, and RedLock, one of those was acquired by Palo Alto Networks. So being able to check your security controls, much like some of the controls that we were reviewing earlier around access control and permissions and things like that. And then there's Cloud Workload Platform Management, or CW PP, which is really looking at overall cloud workloads and container based workloads, things that really can continually assess and mitigate the risk and vulnerability profile of some of those workloads. Aqua, Twistlock, StackLock, those are some of the vendors in that space. And what we always found oddly missing in all of this again, is you could invest millions and millions of dollars in all of these technologies, and ultimately, you're still not protecting the data which is where you have to admit that an attacker is going to get to if they are anywhere in your environment and able to move laterally.

Harold Byun:

And so, we just feel that is something that represents a fundamental shift. And when you again, go back to what is my framework for assessing risk, and when I assess that risk, where do I see those risks coming from, there's a big gaping hole here. And then when you apply that to what security controls are available to me to potentially address those risks, well, if you haven't even put this in your wheelhouse, then you're certainly not going to look any further in your basis for how you're approaching the threat model and risk mitigation measures is a little bit off kilter. This is another view of what we would call a cloud data protection platform that represents some of the different foundational components that you can expect in each of these environments. The baseline where I think everybody operates from is Encryption At-Rest. I think we've beaten that horse a little bit today in terms of what the efficacy is of such a solution or risk mitigation measure.

Harold Byun:

But as you go up the stack here in terms of what you want from a data access monitoring and visibility standpoint, there's a lot of tooling that all the infrastructure providers have exposed in terms of monitoring and logging of events and access that are available to you. There's things that can be done from a data viewing or exfiltration control mechanism, whether that be some of the masking or the hashing or de identification components that a lot of people are starting to look at. And if you start going up beyond that, you get into more data centric measures of protection. And so, whether that be at the column level or in some cases at the cell level and other deployments or methods of deployment, how do you make that a lot more adaptive from a control standpoint, or role based on identity. And so really getting back to something that I think a lot of people are familiar with which is who was accessing the data with what context and should they be?

Harold Byun:

There's a lot of different ways people say that, and then there's some nascent markets where there's things like secure computation and opaque computing models, that are continuing to emerge. But at the end of the day, people have a ton of data funneling into different environments today in a very

unmanaged fashion. And so there's a strong need for trying to tackle this sooner than later obviously. This is something else that is becoming more prevalent, as well as hold your own key or bring your own key, different types of deployment models. A lot of vendors are moving in this direction in terms of being able to enable your own master key. And basically, when you kill a master key that you have control over, that effectively kills the data in the SAS providers environment. This allows for a multi tenant segmentation of the data within the SAS provider environment still gives you some level of control the fact that you hold the master key. And ultimately, if you kill that key, it burns the data inside the SAS vendor or cloud infrastructure providers environment, giving you a sense of data replication which is often a requirement within some of the modern data privacy regulations that are out there today.

Harold Byun:

Again, encourage you all to ask questions in the chat as we go. It's the last tail end, and then I'm going to get into a quick demo. And happy to answer any additional questions that people have. Mitigating the risks. Again, going back to the fundamental question, who can see what data under what conditions? And so how are you identifying users? What is the method for that session identifiers and contexts around the user or group, or the application profile? In many cases, if you're talking about containers, and how they may be accessing an environment, what is that context and attribution? How else can you enrich that context or condition information so that you can make a better decision? And then ultimately, what is the data class or data type that they're asking for in terms of being able to make a request for that data and should it be fulfilled?

Harold Byun:

And while this may sound like it is very heavyweight, there are a lot of methods that are coming that are already in market and there are a lot of ways that people are extending metadata or logical classifications to facilitate a more automated way to enable this type of scenario. These are some common methods for de identification. Again, the slides are going to be made available to you, tokenization format preserving encryption very well known, full blown data encryption, data value encryption. There's application level encryption, column level encryption within that dynamic data masking. Role-based data masking is something that we've seen more prevalently lately as your SQL has implemented a role-based data masking capability as at Snowflake. So there are more and more vendors that are looking at facilitating these dynamically presented views to a user based on who you are and what your context is, right? Advanced Encryption or advanced computation, something that we made mention of a couple slides ago. A very opaque computation method that allows you to operate on de identified data to fulfill machine learning and AI use cases without ever decrypting any of the underlying values.

Harold Byun:

Given the nature of things like COVID, and the desire to share information quickly, that may be sensitive, there's definitely been a pretty good uptick and interest in this type of space. More broadly, I think it's qualified as privacy preserving analytics or privacy enhanced computation. And there's a number of upstarts in this space that are plying in that area for secure analytics. When we go back to this fundamental premise data in the clear when you look at just what I would call blunt force encryption approaches, physical disk encryption or database encryption methods, one of the major things that you want to look at is how you can actually go beyond this to a more data centric method. And these are the de identification techniques that are even called out in that NIST SP publication in the most recent

revision. And so this is an example of clear text data encrypted at the field level and being able to do that.

Harold Byun:

In this case, the data is not available in memory, the data is not available on the logs and clear text, privileged users do not get access to the data. And so it does mitigate some of the risks that we were talking about and inevitably, an attacker moving laterally in your environment, if you believe that you will eventually be compromised which would be a wise assumption for any security practitioner today. If you believe that, then the attacker will get to your environment. And if they're moving laterally, this is what they'll see. This is another variant of this, the bottom half is a clear text data model, the top half is a format preserving encryption scheme. There are other techniques that are available, you can see for example, in the email column, you can probably barely make it out, but there's an app sign and there's a period amongst all the randomized text. And so for any application that's doing application validation, those are things that would still be considered valid strings even though they obviously don't look like the clear text in the lower half.

Harold Byun:

Social Security numbers is another example where format is retained. The credit card numbers or passing credit card loan checks and validation checks in the encrypted mode in the top half of the screen. The dates are all invalid dates, but they pass data validation checks as a valid date for birth date, or death date. There's a number of techniques that are available to implement a more secure mechanism that is data centric without necessarily having some of the more intrusive impacts on an application that people generally are wary of. And then, this is another just high level architecture of how data is becoming incredibly more distributed across a number of different footprints. And so, this is an example of an on premise data structure pipelining to cloud and basically hitting an S3 bucket or it could be Azure Blob storage, and then being exposed via some cataloging mechanism and consumed into alternate analytics footprints, whether that be MapReduce or Snowflake, or Hadoop or some other analytic solution. And so, the move towards this type of data pipelining notion is only accelerating. And you see it with snowflake, and whether we could argue all day, whether the numbers are appropriate valuations or whatever, but the net-net of it is that people are moving towards these warehousing and analytics scenario solutions and droves.

Harold Byun:

And it is only going to increase the data distribution and the data footprint that's exposed. An alternate way of doing this aside from just implementing the data centric methods that we're talking about, is also to look at this as an end to end access channel. And so, not to date myself, but with my experience obviously, I'm sure some of you, I won't say all of you dates back to when a lot of stuff was not SSL enabled, or TLS protected, and so that was a big deal. And ultimately, that was an access channel much like the regulations and the publications that we were discussing earlier suggest being able to secure transmission over the wire with the appropriate encryption algorithm. We're a big believer that today's modern access channel is much more distributed as we covered in some of the earlier threat modeling scenarios that we were discussing.

Harold Byun:

And that is an end to end channel that needs to be secured more appropriately. And so it goes beyond just the data centric protection, it goes into things like the dynamic data masking in conjunction with a role-based access control that really takes a look at who is accessing the information, do they have the appropriate roles and authorization? And if not, how do we contain what they can actually see out of a given data structure. And it is this type of end to end approach that is really going to more thoroughly mitigate the risk of data breaches versus Well, a, versus not doing anything which is what a lot of people do, and then, b, implementing the wrong type of security measure control which is also what a lot of people are doing, and really looking at ways that you can further extend this to mitigate the risk going forward.

Harold Byun:

In summary, and then I will jump into the demo real quick, we believe that you should be reevaluating the modern day hacking approaches and how data leaks are actually happening, measuring that against your data. There's definitely a lot of education that needs to happen amongst the key stakeholders in your organization and also amongst your peers. I can easily tell you that 60 to 70% of the people we talk to, after we talk about data centric security and different approaches, we'll step back after a 30 minute conversation and say, What's the difference between physical disk encryption and TDE?" And that's a very common response. It's gone down over the last couple years, But there's a lot of education that needs to happen in terms of what is the threat that you're mitigating against, and how can you enable your stakeholders to better comprehend what that threat is and what needs to be done.

Harold Byun:

And we've spoken with a lot of CSOs, who have board level discussions, if you can imagine having that with a board member around some of the differences of data protection implementations, whether that be disk encryption, or database level, or an alternate mode of data protection, and ultimately, that's going to make an organization much more resilient. Cuttable 2 or 3, just move beyond the checkbox compliance measures, I think that's a big challenge for a lot of people, checkboxes are easy, it's very easy to check the box and say, "Yes, I feel better about myself today because I feel more secure." And that may make you feel better to a certain degree, but the reality is that it doesn't necessarily do anything to mitigate the data risk. And so, you're really going to measure that as how what the security mandate you want in your organization, and how you want that to actually be operationalized.

Harold Byun:

And then the last point is really the release cycles and development models are all accelerating, the big move to shift left and DevOps. And so, you really want to look at moving into operational models that minimize the impact of DevOps in the business. And that's hard to do as a security practitioner. I get it, I lived it for a long time, but there are places where there are win wins, and I would encourage you to research those. For those of you who are not interested in the demo, I'll come back to the slide. Again, my contact info is here if you have any questions, but let me jump into the demo now really quickly. And again, encourage you to ask any questions in the chat window.

Harold Byun:

Get over here really quick, is asking me to share, go here and share. I believe you can see this now. I should help you can. So this is our UI, and what it does is, we call it the Baffle manager, and I'm still logged in. And what it does is it establishes this data protection layer which interfaces with a key

management solution and marries that information with your data structure. And then, allows us to encrypt data on the fly without ever changing any application code. And so we can do this in line via that data pipeline diagram that I was talking about in terms of piping data to the cloud. We can also do it for any in place application. And so, the first example and I'm just going to show you this database which is in cloud, and I'm just going to invalidate and reconnect, some times these connections get stale. And I will basically run this lookup on this set of tables, and I established this table. And you'll see it's called insecure data, but it's got a bunch of operational information.

Harold Byun:

Forget the data, but the point is that the data is in the clear, and I'm direct to the database right now. Now, if I go to this instance which is this same field and I run a same look up, not on that, I think it's called insecure data, and then I invalidate, reconnect, again just to make sure it's not stale. And I run this query, and you see I get the same data set, and it's in the clear. What I'm going to do here is I'm going to show you what that application looks like. I have this application, and I'm going to choose to encrypt. And if I go to table picker, I have this table called insecure data. And I can basically just say, I want to encrypt a bunch of data, do a couple dates. And so we do have a bunch of different encryption mechanisms. We have a bunch of different formats as well that are available for us. And then I can basically create a key mapping from a key repository. And so, we support all these HSMs and cloud key managers, and if I hit next and deploy and migrate, basically, what's going to happen is, this is going to kick off the job.

Harold Byun:

And effectively what we're doing is we are retrieving or fetching a key in a key store. We're decrypting it using a master key and we're mapping that key ID two or three to the data columns that I had selected in the prior view. And we're basically encrypting this on the flyer in place. And so this isn't a huge data set, I think its 1.1, 1.2 million rows, should take just a few more seconds, but basically, we're kicking off this job, it's going to fetch those keys, create that mapping and encrypt the data. So should just be another moment or so here. And there goes, almost done. For those of you who are interested in performance numbers Yes, look there's no free lunch as it relates to encryption, just to answer this question, we've definitely greatly minimized the overhead and there's other things that we've done to significantly accelerate or the way traffic is passed through us.

Harold Byun:

We are in production globally with two of the largest banks in the United States through a SAS provider, and that's an environment with over 10 billion data records. And they've measured us at one to two milliseconds of overhead in terms of data migrations like I've just done. We've clocked roughly 80 million records an hour on a single instance. You can obviously parallelize the effort, but it's fairly optimal from a data performance perspective. Now, if I go back to the direct access method, this was the direct one, and I refresh, you'll see that we've encrypted a certain set of columns. And so, this is again protecting you against that lateral attack.

Harold Byun:

And then what we do is we establish this transparent layer which is this query model. And I'm encrypting this, and you can see that encrypting that all the data is in the clear. And so, we're basically performing that encrypt-decrypt function, but we can also do it with access control as well. And so, the way that it

looks with access control is if I go to this particular environment and do something like a lookup on this table, you can see that this name column is encrypted. A lot of these look encrypted, but they're not, they're just test data, so don't get confused by that. But this is an encrypted column, I know.

Harold Byun:

And so what happens is when I go through this Baffle shields component, and I run the same lookup, you can see that we're masking the data now, so that nobody can view it because we've done a global lookup on that actual table. But if I'm able to pass IAM information or data owner information, I can selectively decrypt certain rows within this data structure and passing a different identifier, I get different rows. And so the point being here is that all of the data is encrypted underneath. And what we're using is a context switch with consuming and identity and basically allowing for a user to selectively only see the data records that they should be able to see.

Harold Byun:

That's the quick and dirty demo. I see that there's some additional questions here. For those of you that aren't interested in questions, there's more downloads available at our website. And let me see, what databases or data structures Do you support? We support Microsoft SQL, MySQL, MariaDB, Postgres, Redshift, Snowflake, Amazon S3, Azure Blob Storage, Azure SQL, GCP databases, all the databases they have in their platform. Those are the major data structures. We also have an API, if we don't cover something, always leverage that.

Harold Byun:

Our containers easier to hack? I think that containers are equally as easy to hack as any other systems. So you just have to look at your overall security posture. Are you guys a SAP service? No, we are not a SAP service. We like to say that we don't want your data and we don't want your keys. So we give you software so that you can establish your own data protection layer and service. And we've had a lot of great success operationalizing with large scale customers.

Harold Byun:

Comments on the number one security risk social engineering? Yeah, I completely agree, the one thing that I would say is with social engineering is that oftentimes that leads to a compromised credential in some form. And when you're talking about somebody with a compromised credential and the overall blast radius of that credential, combining that with an access control mechanism, you can basically enable a dynamic entitlements mechanism or method that allows people with context only see mass data. So another way of thinking about that is, Yes, social engineering is real, and Yes, people are going to get around all the time.

Harold Byun:

But there are a number of ways that you can also detect some of that behavior. And as you are able to detect that behavior, there's ways to more automatically trigger restricted access to the actual data structures, versus lying in wait or looking at ways that you can catch up to all these users clicking on different things or getting owned in different ways. Very real threat. I think it's a valid comment. I just think that there's ways to more quickly automate within the security response ecosystem that you've built out to tie this together and enable adaptive data security.

Harold Byun:

Last question, unless there's more. Other specific key managers that you support psychotic. We support psychotic Azure core, Azure Key Vault, Secrets managers, AWS secrets manager, all the major HSMs, we have an integration with TALAS and Keysecure, KMS. When we use industry standard protocols to support the retrieval of keys, it's a two tier key management hierarchy with a master key that encrypts any data encryption key. So you kill the master key and you basically have killed all the data encryption keys.

Harold Byun:

Last question. Okay, for the row based contextual scenario, are you decrypting inline, is the decryption outside the HSM boundary? Yes, we are outside. Well, the decryption of the deck happens in the HSM. And the decrypted deck is what is actually held to perform the inline decryption. So nothing is actually leaving the deck from a master key perspective and the operations are all held within the HSM, which is the way that it should be happening. And that's occurring via PKCS 11 Library, but the actual act of the data being decrypted is happening within the application flow. We're not passing data values into the HSM or anything like that, that would not perform or scale.

Harold Byun:

But within the constraints of the security context of the HSM or the boundary that you're referencing, the master key never leaves the HSM and the decryption of the DK is happening inside the HSM. All righty, I hope that this was interesting and helpful. And we're happy to talk in more detail anytime if you've got other questions. And we'd love to engage, and we try to be just generally pretty straight up in terms of our approach to security and what we can do and what we can't do, and what we think might be the best fit from a security and threat modeling perspective. I hope you found this helpful. Really appreciate you taking time out and have a great day. Thank you.