Harold Byun:

Welcome and thank you for joining us today. There are several registrants still signing in so we will pause for the next minute or so to allow everyone to get settled in. Thank you for your patience.

Harold Byun:

Hi. Good morning, good afternoon, good evening. Thank you for taking the time to join us today. My name is Harold Byun. I head off project management for Baffle and we appreciate you all joining us for the webinar, ensuring a data privacy for SaaS. Implementing simplified record level encryption.

Harold Byun:

So today we are going to be reviewing just a number of topics. Hope to get in and out of here within roughly a half hour or so. The agenda for today, we're going to review some common security and operational challenges in SaaS, which we call shared data or co-mingled data environments. We'll look at some scenarios that go beyond just SaaS usage or SaaS implementation. We'll cover BYOK architectural considerations and how to establish BYOK as a service. And then we'll review simplified record level encryption with a live demonstration of NRLE solution as well as covering data shredding and the quote unquote "right to be forgotten" which seems to be a bit of a cash range with some privacy regulations today. And we'll finish with some Q and A. So, glad you could take the time and we'll jump right into it.

Harold Byun:

You know, in terms of some SaaS security challenges, that we're going to cover today, the first stuff that we're going to cover is just some definitions to level stat to cover some of the acronyms. I'm sure most of you are familiar with them, but just to ensure that we're all on a level playing field.

Harold Byun:

I'll pretty much gloss over this one. SaaS, software as a service. On demand hosted software delivered as a subscription service.

Harold Byun:

BYOK, commonly known as bring your own key, as in bring your own encryption key to encrypt your data.

Harold Byun:

Record level encryption or RLE, or row level encryption versus field or column level encryption versus data or file encryption. This is a common misconception across a number of folks that we talk to in the market and so I will cover this in a little bit more detail in a couple slides. We won't spend a ton of time on it. There's definitely some more information on our website as well.

Harold Byun:

The customer. I'd like to extend the definition of the customer. In many cases that you, who are all joining us or some of you may be vendors or SaaS vendors, and in many ways the customer is you or

your company, but it really extends to this notion of a subscriber and ultimately that of a data owner or what we call frequently internally here at Baffle, an entity. And so I think that you can kind of look at those as all equivalent and as you kind of look at some of the scenarios that we're going to cover today, it extends into a multitude of other scenarios that go beyond just a SaaS data protection model.

Harold Byun:

In terms of privacy regulations and the context of challenges around SaaS today, you'll see that the... obviously privacy regulations are continuing to emerge. Obviously GDPR was the prominent one. We have CCPA which is now actually just starting to go live in 2020 but all of these measures around data privacy continue to emerge. And you can even see it in the consumer trends in terms of a lot of the common consumer oriented company messaging and advertising their privacy strengths and advantages in the market. We fully expect another 15 to 20 states to pass privacy regulation as well as individual countries around the world, so it isn't as if this is going to be going away any time soon.

Harold Byun:

When we look at the other notion of challenges and SaaS providers, there's really kind of a security and operations perspective. And I hate to paint it as an us versus them scenario but I think just nothing's cut and dry or black and white in terms of these types of comparisons but in general you have two different perspectives from an operational and security model. And on the SaaS vendor, in terms of their wants and don't wants, they obviously want to drive a consistent service delivery model, consistent measure of application delivery that is cost effective and operationally efficient. That and they want to release on a continuous basis and a faster fashion, in a highly performant manner.

Harold Byun:

And what they don't want to introduce into their environment as it relates to security is complexity in the operation, any of these type of one-off solutions, and certainly don't want to head down the path of dedicated instances, and we'll cover that architectural model as well. Any types of architectural changes and this notion of an inconsistent support model which often comes out of a dedicated instance model.

Harold Byun:

In terms of the consumer or data owner side, when we really look at this, your business is looking for an easy to consume app, some performance guarantees, and obviously more features. And from the security perspective and the Cloud security architects or Cloud architecture teams as well as security and risk in general, they're looking for security guarantees and we often see that translate into this notion of control or ownership of some form of a controlling mechanism or a right to replication and obviously in the emerging privacy regulation that we're seeing, a right to shred or be forgotten.

Harold Byun:

And ultimately what consumers or customers or data owners don't really care about is, in many ways a lot of the enterprise costumers don't want co-mingled data stores. They could care less whether or not the instance is dedicated or not, unless you're a vendor charging more money for that. And they obviously don't want any performance degradation and ultimately they want to minimize any risk to data and the business.

Harold Byun:

So how does this actually translate into the operating and security model that we see for different SaaS providers? Well what we see is people running the spectrum from what we call single tenants or multi-tenants, so a single tenant, you have the dedicated application tier and a dedicated database on a per data owner basis. There's really just a managed hosted environment and that often maybe remarketed SaaS, some people call it fake SaaS, but at the end of the day you at least have a single tenant instance and there isn't any co-mingling of data in this model. Another variant of this is what we call single instance which is multiple data owners or subscribers, single app tier and a dedicated database per data owner.

Harold Byun:

And then we also have this notion of a multi-tenant environment which is a single app tier with a single database for all data owners and ultimately the data store is co-mingled with some logical separation and I circled it in red because it's just so dangerous. No, I'm being sarcastic there. I've circled it in red because that's probably what we're mostly going to focus on today in terms of some of the challenges in implementing a data segmentation model within this type of multi-tenant scenario that goes beyond just logical separation.

Harold Byun:

So when we look at securing multi-tenant SaaS, there's a number of challenges here, one being can you consume customer owned key material and how could you actually implement this? Once you actually make the decision to do that, how do you actually modify the application to consume the key material? How do you get it from external sources and do you do this on every single release of your application? If you're a CICD deployment shop you're ultimately releasing several times a day and you have to ensure that this capability is wired into each rev of your application. And then what happens when you release new services or applications to market? And it becomes this an increasing overhead on your operations as you release new applications and services. And then ultimately, how do you take this a step further? Take that key material and map it to data in a shared or co-mingled environment in a multi-tenant model? And then the last question is, as SaaS vendors or providers, even as an IT service provider, wouldn't you rather focus your development efforts on other features and functionality verus wiring up application, after application to do this.

Harold Byun:

We think that there's a better model and we'll cover that obviously very shortly. Some of the scenarios that we see that's applying to beyond SaaS are a few different scenarios. I'm actually going to shift into a screen share here so we can actually just cover some of these, or I think I'm going to. Give me one second here.

Harold Byun:

And with this, different models that we see, this is one model that we're going to cover so when we look at different data classification, so you can see here that we have this multi-tenant SaaS application and there's a method to not only map data to a specific data owner, but you could shift that and say it could be almost any metadata or data classification. And so if we wanted to segment data in an environment that may have different data classifications, there's ways that we can implement a record level encryption scenario for shared or co-mingled data to implement this type of internal fire walling. Could

be very useful for joint ventures, very usually when you're dealing with multiple third parties. Again accessing a common data store who need to basically access data.

Harold Byun:

Consumer right of revocation is another scenario so again, alluding to some of the privacy regulation that we're seeing so the ability for an organization to support billions of key mappings at the record level, the ability to enable the right to could ultimately extend down to a consumer, as well as selective data masking for different owners based on an entitlement. And the way we actually see this being implemented is cross scenarios where healthcare providers may have a scenario where someone wants their data to be revoked or it could be somebody in biotech or retailer with consumer information and so these are methods that could be extended beyond a typical SaaS scenario that could also provide a level of data privacy.

Harold Byun:

And then the last one is really around third party data access control. It's kind of a step forward on the consumer model or the different data classification model. In many scenarios, a lot of organizations are dealing with hundreds if not thousands or in large enterprise, tens of thousands of third party suppliers and vendors, and often times data is shared pretty much left and right across those vendors once they fill out the 150 page questionnaire or their sig and once that's completed, the data is kind of off in the wild and another approach to this could be to facilitate a better model of data sharing and data privacy using this type of, again, security model that locks down the shared or co-mingled data store.

Harold Byun:

So moving ahead, how do we actually start implementing this when we look at multiple data owners in a SaaS or shared data environment? And so there's obviously the bring your own key considerations and how do you actually make this work? One way is to go the long road which is what we were kind of enumerating earlier which is this path, which is you figure out how to rewire you app, you embed some kind of encryption capability or SDK. You figure out a key exchange for the respective key stores. You find a way to talk to those respective key stores and then you implement this on a per record basis using some kind of identifying factor to support this multi-tenant encryption model. It's very very heavy weight. It is not easy. It is not a slam dunk by any means to implement and we think that there's an easier path forward for people.

Harold Byun:

So one way of doing this is you could use something like a customer owned customer master key support model. This is one model that there's a link to some of the Amazon or AWS docks where there is a method to use a mapping to facilitate a customer controlled or customer owned SMK that is used for encryption in a given environment for some type of SaaS provider. And that doesn't necessarily map it to the record level but at the very least, at least you have a customer owned key model.

Harold Byun:

Another variant of this is what we call the customer generated key model. The customer key generated model is where the customer one or customer n in this example is providing some form of generated key material which is utilized in conjunction to create a customer master key that is specific to the customer but is ultimately held in a SaaS provider key store. A lot of people don't like this model for

whatever reason and it makes sense, people want ultimately control over the key material and the right to revoke it, and you lose some of the right to revoke here. And it just isn't I think tangible. It's not the best word for it but people want a key that they can more or less quote unquote "touch and feel" and control and revoke and so you don't necessarily get that in this model versus this model.

Harold Byun:

So how do you ultimately give data owners this control that they ultimately want and so the approach that we've taken is ultimately providing what we call a simplified record level encryption solution. And so I'm not going to get too much into the details of Baffle but at the heart of it, one of the major components that Baffle has in its centric protection model is a component called the BaffleShield. It is a TCP wire protocol reverse proxy for sequel and no sequel traffic, and it has the ability to seamlessly encrypt and decrypt data in structured data environments or no sequel environments, by effectively maintaining a privacy scheme or shadow information scheme, so the application doesn't even know that this exists. Doesn't require any application code modification. You're not engaging in developers. It becomes a data abstraction control layer that also performs an encrypt and decrypt function.

Harold Byun:

What we did with this component is we extended it to record level encryption capabilities, so what we're able to do is key off of a client or an entity ID or a data owner going back to our notion of who a customer is, or it could be a data classification and we're able to extend that same encryption model not just at the column level but also at the record level scaling to in theory billions of records. To date we've done in the range of hundreds of thousands of different keys in environments exceeding 10 billion records, so that is fully integrated into a CICD automated deployment pipeline in an environment with over 10 billion data records. So it is something where we've driven a lot of scale and performance and we're able to again do that at the record level. In theory our model should scale to the billions, we obviously have not to date done that.

Harold Byun:

In terms of the Baffle management overview, really we can support integration with virtually any key manager or key management service. We support industry standard protocols, so KMIP, PKCS#11 libraries for HSMs, and REST APIs for Cloud key managers like Cloud HSM or hazard key vault or AWS KMS. We use encryption key material to encrypt data at both the column or record level. We can be deployed with modifying any tier application tier code and we obviously again support multiple CMKs across these data owners at the record level to support multi-tenant SaaS providers.

Harold Byun:

This says operating and AWS, it really doesn't matter where you operate, if you're a multi-tenant provider and you have customers that are mandating a higher level of data privacy or segmentation for their data, we can deliver that for you in your own hosted environment without you incurring the cost of the application code modifications and rewriring your app or performing architectural overhauls. We can do it at scale and you're able to deliver this in your own hosted environment as well.

Harold Byun:

The way we do this is we basically facilitate a model where we can consume key material and that BaffleShield component creates this mapping and performs that encrypt and decrypt function on both

the column and the record level. The SMPCs servelets are not something that we're going to talk about today. We have a multiparty compute implementation that allows for secure computation and mathematical operations on encrypted data so that's another level of functionality that we can offer to perform operations on encrypted data and minimize any impact or adverse impact that encryption may have on your application functionality. So supporting sort and wildcard search for example, but that's out of scope for this discussion today. Just didn't have a chance to modify this slide.

Harold Byun:

This is just an example of some of the UIs, so this is one example of an integration with a key management store. We basically create a master key and a data encryption key hierarchy where the decks are encrypted with a master key and obviously we can support multiple master keys and can rotate either master keys or decks. And we can do this across again multiple key providers.

Harold Byun:

This is just more of a detailed view of what we're doing with customer manage keys in native US. What we do is we uniquely identify respective customer manage keys by ARN and so the key can be generated and owned and disabled by the customer and it is referenced by the ARN which is a unique pointer across Amazon or an Amazon resource name and so we are able to basically consumer the customer controlled key and if a customer disables or revokes the key effectively the ability to decrypt any of the data is gone and so that data effectively is shredded supporting this right to be forgotten.

Harold Byun:

This is just a variant, the only variant of this and the other slide is basically we've popped our BaffleShield into the SaaS provider. This is where we have a number of organizations again that have implemented us in their environment versus taking the long toll approach of modifying their application and architecture or going down the dedicated instance route. And in this model the BaffleShield, which is software, runs inside the SaaS provider environment and ultimately gets deployed as part of their own operational process so we're able to do that very very efficiently and seamlessly and in some cases our providers have just completely wired us up in a containerized Cloud-native environment and deployed us as part of their CICD pipeline.

Harold Byun:

So let's kind of get into what this actually looks like. This will be in a fairly quick rudimentary demo, I will try to make it a little bit more interesting, this is my nifty little console that I like to use. It's kind of one of the TVs that I grew up with, and what I'm going to do is I'm going to get into this particular instance and I will... let's see. And over here and I'm going to log into an environment. I think you'll see that when I get into this environment what I'll do is just execute a base level query and you'll see that this query is looking for loan number and property ID. And when we run this query we can see that we get some property IDs but the loan number in this particular case is actually encrypted garbage. I could run this a different way and render it as hex so you can see it more clearly. I hope you can see this actually and you'll see that column is encrypted, we're just obviously displaying it in a hex format.

Harold Byun:

Now if I go to the BaffleShield, and log in here. And run that same query, you'll see that we're decrypting the loan number. So this is the encrypt decrypt function that the BaffleShield actually provides you. And

what I'm going to simulate here is disabling the key. Obviously somebody would be normally doing this from their HSM or whatever key management solution they're using, I'm going to just simulate it here.

Harold Byun:

And when we go back in and run this same query, you can see that there are certain records that we've actually masked. And the reason that we've masked this is because this data isn't just encrypted on the back end, it's encrypted again on a per record basis that is tied to a given data owner and based on that respective data owner, disabling or deleting their key or revoking it from the provider, we're effectively no longer able to render that data in the clear. And so that is kind of the customer owned or bring your own key service that you get via effectively an off the shelf offering from us to facilitate this type of implementation and operating model versus building it your own which is what we were covering in the prior slides.

Harold Byun:

So there's a number of other variants around this where we can employ data exfiltration controls. We have a port for formatted masking so things need to be rendered in a certain format like a phone number or an SSM or a specific account number format. We can support that easily as well, but ultimately we do this so the application doesn't get fed garbage and crash which can often happen if encrypted data isn't handled properly.

Harold Byun:

So this is actually a... so this is what effectively we can provide from a simplified record level encryption solution for SaaS, as I mentioned earlier we're deployed with several large scale SaaS providers. One of whom is over 10 billion records in deployment. This is just kind of their quote, I'll summarize, you can always read it, and we can make this deck available for you, but they wanted to deploy a bring your own key model effectively for their customer base with revocation rights and use those keys. And the joint solution with Baffle their SaaS service, plus Baffle didn't require any large scale architectural overhauls or application changes or dedicated database per tenant. And as a result, the development time was instead being spent on adding even higher value enhancements. SO very happy and pleased with some of the partnerships that we've been able to establish with different SaaS providers in market. This one is serving the bulk of the fortune five hundred in terms of their enterprise offering.

Harold Byun:

Data shredding and the right to be forgotten. This is effectively what I was showing you in that demo, the ability for us to not render data where we don't have a key present is kind of the bulk of the offering in terms of a simplified solution. So hopefully you're getting a sense of how this technology could be extended into some of the other scenarios beyond SaaS that we were recovering earlier in the presentation.

Harold Byun:

And then overall, what Baffle has really kind of delivered is an aesthetic capabilities to the market that we believe, or that we've called, a Cloud data protection platform, and it encompasses the field level encryption, the record level encryption which extends into different data entitlements mechanisms, the advanced encryption we didn't cover, which is the multiparty compute model that we support which can again support operations on encrypted data. We have an exfiltration control mechanism to minimize or

mitigate the risk of data breaches or mass exfiltration. And we have a data access monitoring product that's coming to market later this year as well for overall monitoring, alerting, and anomaly detection.

Harold Byun:

So those are all things that are being extended into kind of our comprehensive suite of the product portfolio which includes basically a Baffle manager which is the admin UI, the BaffleShield which is what we were just demonstrating in terms of the encrypt and decrypt functionality, to scan an invisible data abstraction layer that sits below the application tier, and then our secure multiparty compute implementation which was basically the first to market to support wildcard search on ADS encrypted data without every decrypting the data in memory in use or in process.

Harold Byun:

So kind of the bulk of the, again, trying to get in and out of here in roughly half an hour, we'll open it up for questions. I'll leave this up for folks that are interested or actually maybe I'll switch back to the other slide deck so that you guys can basically look at that while we take some questions from the audience. So let me kind of go ahead there. So feel free to use the chat box or I think there's... is there a question box? Okay. So yeah, use the question box. I don't have the resources up there but you can obviously go to our website. So I will kind of see what we have from a question standpoint here.

Harold Byun:

The first question is how do you integrate with key managers and which key managers do you integrate with? So as I mentioned earlier we support all the standard industry protocols for integrating with key management solutions so KMIP client, one not one or higher, PKCS#11 libraries is kind of the industry standard protocol for HSM integration. We can support Cloud HSMs. Could be a war metric DSM#1 or we have a partnership with Thales which I think owns Gemalto which owns SafeNet and so half the encryption world at this point. So we've partnered with them very heavily and then all the Cloud key managers are via rest API solutions so that's basically how we integrate with key managers. We support what is commonly known as a two tier hierarchy so again a master key and a respective data encryption key, what we do is we encrypt the decks with the master key and either can be rotated and we can also support multiple versions of keys.

Harold Byun:

How many keys can the solution support? I think I touched on this earlier. It's roughly theoretically in the billions. You might've caught it when we were rendering some of the hex rendering of the encrypted data in the quick demo, there is so metadata that is prepended to the encrypted values and in that we actually track basically the data owner identification as well as the key version information and in theory that should scale to the billions. We've scaled it up to hundreds of thousands at this point but haven't gotten beyond that. We have some healthcare providers in region that are looking in the tens of millions of records for a consumer facing model and there should be nothing to prevent that from operating at that level.

Harold Byun:

Do you support Cloud native environments and containers? Yes, absolutely, so we have deployed a Cloud native solution. We've been dockorized, if that's a word, and incorporated into a cooper '90s framework so very easy deployment. We also have our own Cloud formation template. So we have

application stacks that are available and in an AWS environment you can deploy our solution in roughly three and a half minutes using those templates in an HA pair spread across different availability zones and registered with an elastic load balancer, so we've vastly simplified not just the encryption model and the abstraction but the operating model behind it. So those are some things that are in place.

Harold Byun:

Do you support creating policies for RLE? I think it depends on your definition of policies. What we do is we ultimately operate as an enforcement vehicle, I guess is kind of maybe the best way to think about it right now. So based on your policy, we basically maintain an entity mapping or data owner mapping and a privacy scheme that maps the respective columns and the data owners map to the respective records. The respective data owners could be a data classification. It could be a given policy set or regulatory policy enforcement and we would utilize that to enforce that at the record level.

Harold Byun:

If you're asking if we can enforce things more dynamically based on user context or attribution for a dynamic data entitlements mechanism, those are things that are relatively near turn road map and if you pass SaaS' session ID or some type of user session ID, then we could definitely consume that and enforce a dynamic policy around a user entitlement so it could be, here's a third party vendor that we work with on a regular basis. Today the third party vendor is coming from a new VPC or a new subnet block and is asking for a higher volume of data, based on that we could enforce a specific type of RLE policy that would minimize risk if you view that as a risky scenario. Or the more obvious one is a user that hasn't two factored and is coming from a different country today which may be a compromised credential and asking for a certain amount of data, those are all different types of scenarios so I hope I answered that question for you but if not, feel free to ask again or you can always reach me at the email address on the screen.

Harold Byun:

What is the storage expansion cost? I think this will be the last one unless there are more questions. So our storage expansion cost is a fixed cost of the existing data set for encrypted data. Obviously the bulk of the data is probably still going to be clear text if you're using a more granular column or record level solution, but for the encrypted data there is always a cost to encrypting the data. We can support a format preserving encryption scheme if desired but that format preserving encryption scheme, while it would not have a storage expansion cost, would not be eligible for record level encryption because there's no way to track the metadata on to the data owner is. And so there are alternate ways you could do it but they're nonperformant and they would never scale in the environments that we're running in today. So the storage expansion cost is a fixed cost of storage.

Harold Byun:

If anywhere, it's roughly close to around 50 bytes of a fixed cost and so depending on your data types that you're encrypting, would either a small or a large percentage. But at least it's a known quantity, it's not doubling the storage or it's not going one and a half times. It's 50 bytes and so if you're doing this on long text or free form descriptions or comments or things like that, then the percentage overhead is relatively small. If you're doing this on an integer like H, which has a length of three, then it's going to be high, but it is at least a fixed cost.

Harold Byun:

A couple additional questions, last question is coming in. Your encryption applied to data at rest or at the same time it's getting created? It is as it's being created and inserted and updated and then it is also at rest. So as you're reading and as you're writing data or creating it or generating it and putting it into a data structure, we are encrypting it on the fly, and then it obviously gets encrypted at rest. Once the application is presenting the data, it's always in the clear at the presentation layer unless it is leveraging some of the masking capabilities that we have and you've intentionally decided to mask certain fields from being displayed. But otherwise we're always decrypting at the presentation layer. We also have our own migration utility that allows you to migrate data from point A to point B and encrypt it on the fly.

Harold Byun:

It's not really creation but it's kind of in transit or from point A to point B. And then we have the advanced encryption model which also allows us to run computations on encrypted data which also encrypts the data not only at rest and at creation time but also in use and in memory.

Harold Byun:

And then the last question, you're never too late, but does the Baffle software capability reside on customer computers or networks, or does it require a connection to the Baffle network? That's a great question. So we are software that is deployed in your own network. We don't want your stuff. We don't want your keys, we don't want your data, we want to help you protect your data. You can deploy us on premise. You can deploy us in Cloud. You can deploy us in a hybrid model. You can deploy us in your own VPCs. So that's kind of our model and you can deploy us anywhere.

Harold Byun:

All right, last question. Do you support file level transparent encryption? So if you're asking if we interact with file level transparent encryption or file system encryption, yes, completely seamless to us, that is operating at a level that's below us, so we absolutely would support that. We would be completely, I hate to say that we're oblivious, I can't think of a better word right now, we would not have knowledge of that transparent encryption model, and then we also have some things that we're doing on our own file or object level encryption that are going to be coming to market later this year as well.

Harold Byun:

So I hope that you all found this to be helpful. We try to keep them short and sweet. If you want more information we're more than happy to get on a call and discuss things in deeper technical level discussion and again, happy to do it as time efficiently as you want or happy to spend more time with you if you want as well. And so we can be reached at info@baffle.il. I'm harold@baffle.il or on Twitter and please let us know how we can help you as you kind of do your due diligence and research in this market space and hopefully this was of interest to you.

Harold Byun:

So thank you very much for your time and have a great day.