Humane
On-call

Good afternoon SREcon.

I’d like to talk to you today about a topic that impacts the lives of a lot of folks who work in SRE, technical operations and software development roles.

On-call is a key enabler of the non-stop operations that provides the foundations for most modern businesses.

But it’s not always approached with the thoughtfulness required to take care of the humans involved.

Martin Barry

marty@supine.com

@supine

But first a little introduction…

I’m Martin, an Australian who moved to Germany 14 years ago.

I’ve worked in network and system administration for more than two decades across a variety of companies and industries.

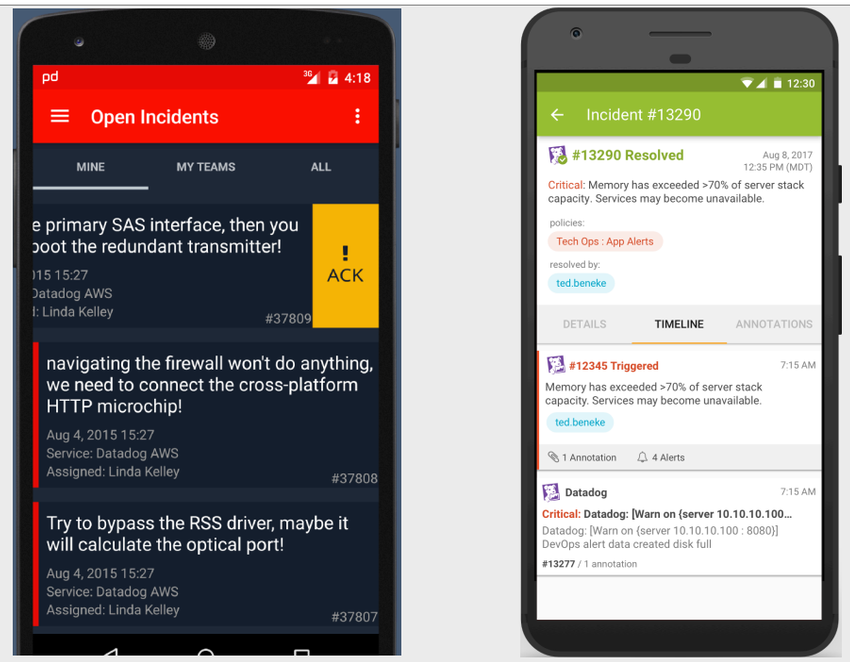
On-call

I’m sure everyone in the room is aware of the value on-call provides, keeping businesses and infrastructure operating 24 hours a day, 7 days a week.

It’s not just for highly visible things like ecommerce and social media, but extends all the way through to critical components of our societies like transport networks and hospital systems.

A quick show of hands…

Who in the room has been, or is currently part of, an on-call rotation?



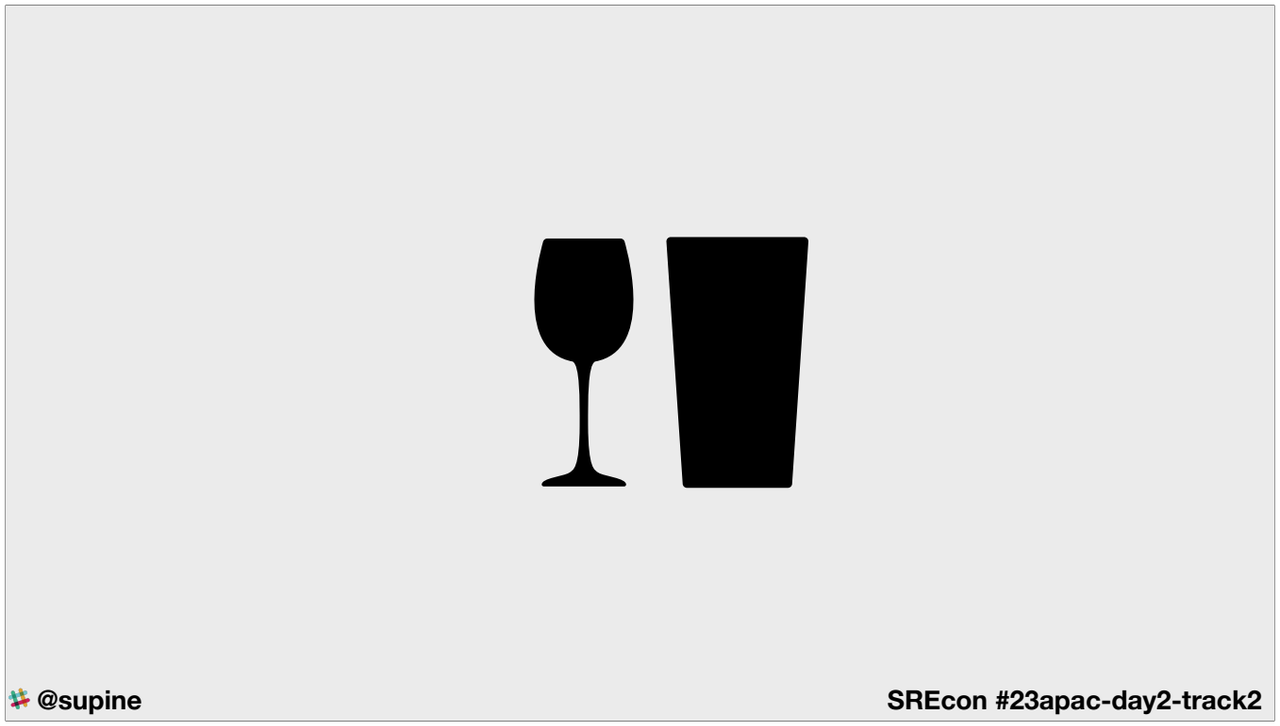
I’m sure many folks with their hands up are familiar with tools like these.

How many of you have memories like this?

What about this one?

And how many of us are this old?

…



Lets keep the questions going…

How many of you make a point of not drinking any alcohol while on-call?

And how many would allow yourself one standard drink while on-call?

Two standard drinks?

More than two?

Many folks choose to abstain from alcohol or moderate their intake while on-call because they know it can reduce their capacity to respond effectively.

Many companies have policies that ask folks to not consume alcohol when on-call for the same reasons.

Slower reaction times

Impaired judgement

Increased risk taking

We know alcohol affects things like our cognitive abilities and response times because it’s been studied.

Lets get another show of hands, whose company has a policy that you should not drink alcohol while on-call?

…

Do we apply that same care when it comes to other things that make us unsuitable to work?

Fatigue

Sleep deprivation

Keep your hands up if your company has guidelines around fatigue management while on-call?

Do they articulate how to recognise tiredness and the appropriate responses?

Difficulty concentrating

Decreased alertness

Lethargy

Like alcohol, we know how fatigue and sleep deprivation affects our cognitive abilities and response times because it’s been studied.

Mental and physical fatigue can make someone unsuitable for work, unable to continue to provide on-call coverage.

Fatigue management

So what can we do to prevent fatigue or manage it once it is a problem?

Got to keep the surveys going so, show of hands…

One person team, on-call all day, every day, 365 days a year…

Who has been part of an on-call “rotation” that looked something like this?

…

It’s an extreme example but it leads nicely into the first suggestion…

Maximise number of people in rotation

One way of preventing fatigue from becoming a problem is to minimise the amount of time any individual is on-call.

This means looking for ways to maximise the number of people in each on-call rotation.

When you are hiring you should look for ways to augment your on-call rotations.

What skills will the role bring?

How does their work location and timezone fit in with the rotation strategy?

Does your job description and interview process make clear to candidates your on-call requirements if they are hired?

Maximise number of people in rotation

If you have a number of small rotations you can look to consolidate them into a single rotation.

This might result in people receiving alerts for problems outside their core skill strengths, so this presumes that there are sufficient runbooks and other assistance so those folks can make at least an initial triage, and have clear escalation paths to get a true subject matter expert involved if required.

Put developers into on-call

Another way to get more people supporting on-call is to widen the scope of roles required to participate in it.

Bringing developers into on-call has been a theme for a few years now but there are many places where nothing has changed and this option is still available.

You might not be able to include them in generic SRE rotations but you can take alerts that can be definitely associated with their running code and send those alerts straight to them, rather than having the SRE on-call triage the alert first.

Minimise the duration of each on-call shift

Another way of trying to address fatigue is to rotate more often, looking to make each on-call shift as short as is reasonably possible.

Week-long, 7 full day shifts are common but that’s an awfully long time to be primary on-call.

If you can, explore alternatives.

Daily rotations might be too frequent but splitting the week into some combination of two, three or four day shifts might make them easier to survive.

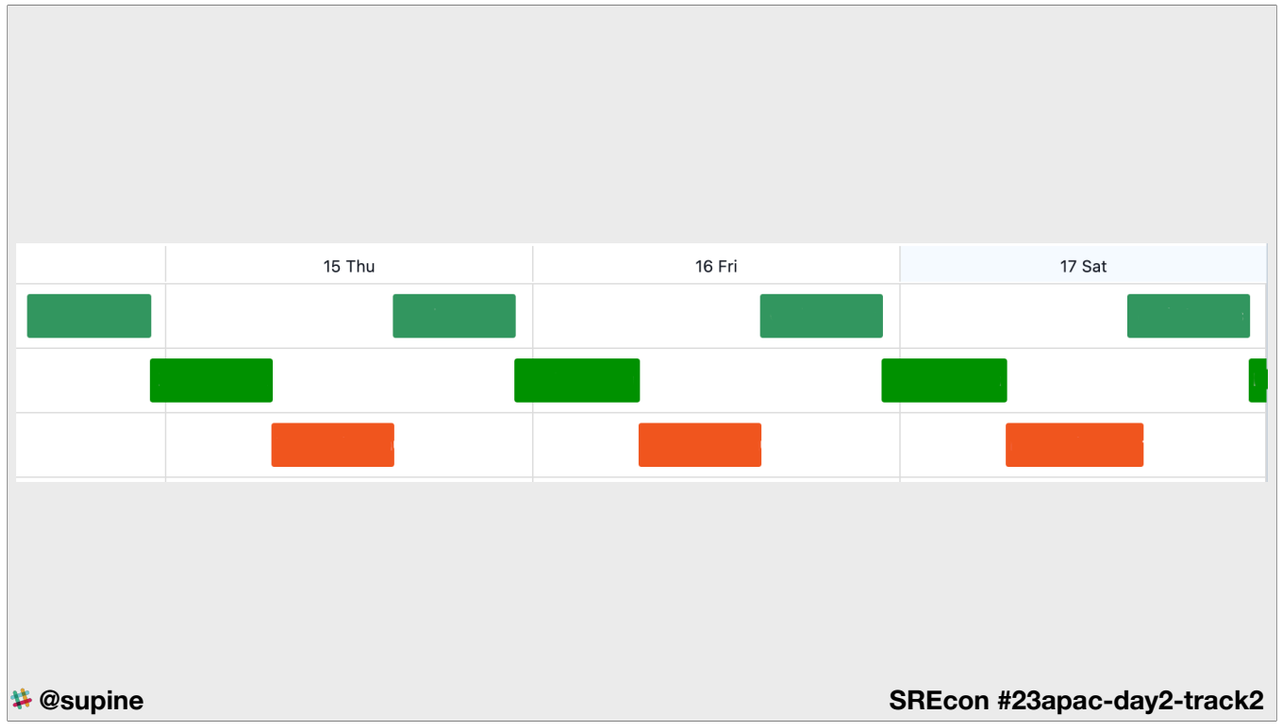
Stop waking people up

There are many good reasons to have a follow-the-sun, global rotation.

Business hours shifts are much easier to staff.

You can tap into talent pools with people who have different backgrounds, skillsets, languages spoken…

But the key benefit for this talk is that we can aim to always send alerts to someone who is already awake, maybe even already at their desk.

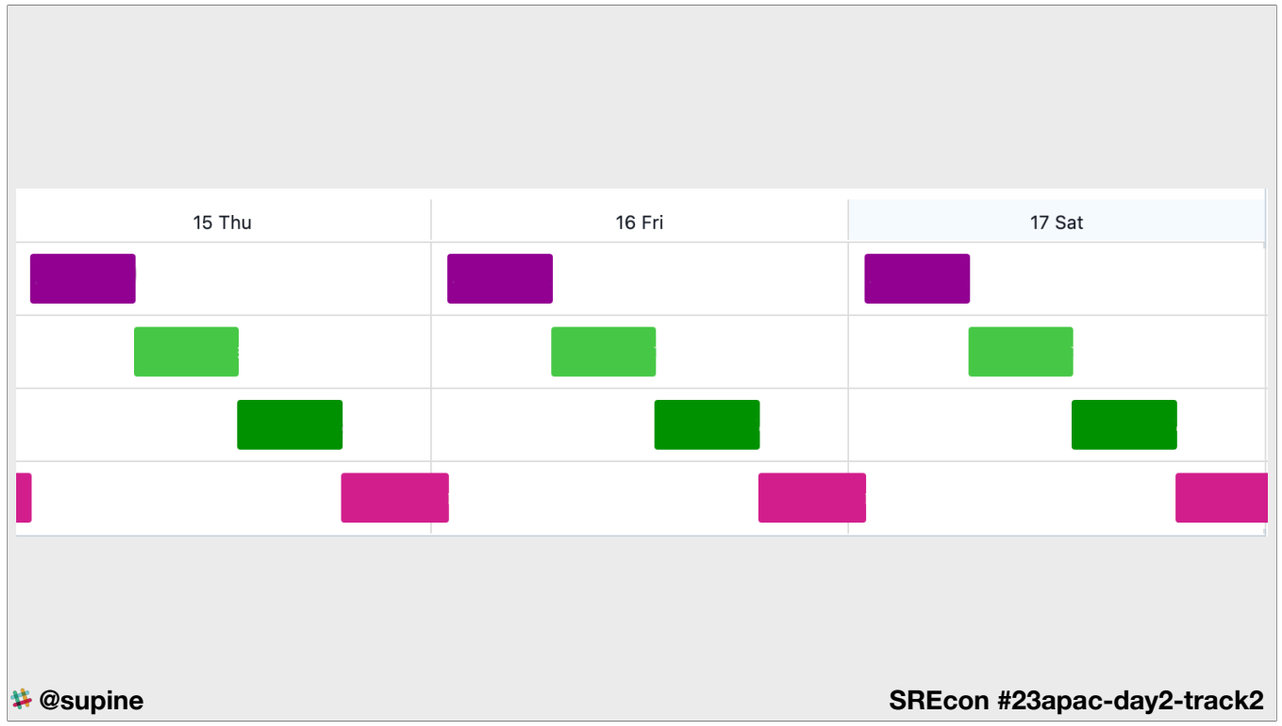


Reasonable follow-the-sun coverage can be created with team members in as little as three different, strategically selected timezones.

One combination that works is having parts of the team in Asia Pacific, Europe and West coast of the Americas.

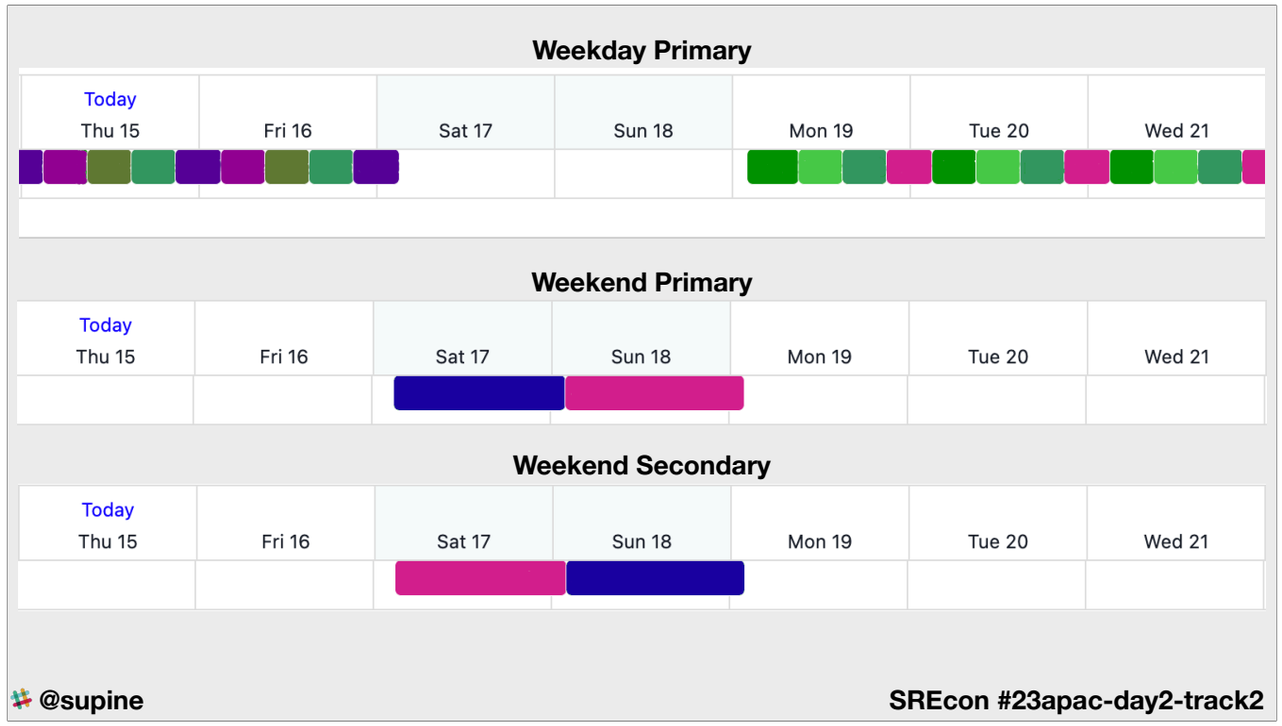
The shifts are 8 hours long, it can be hard to keep them within standard business hours and you need a little flexibility as daylight savings changes can throw everything out of alignment.

But it works.



It gets even better if you have more locations.

The shifts can be made shorter and its easier to keep them within standard business hours for each location.



And you can get quite complex…

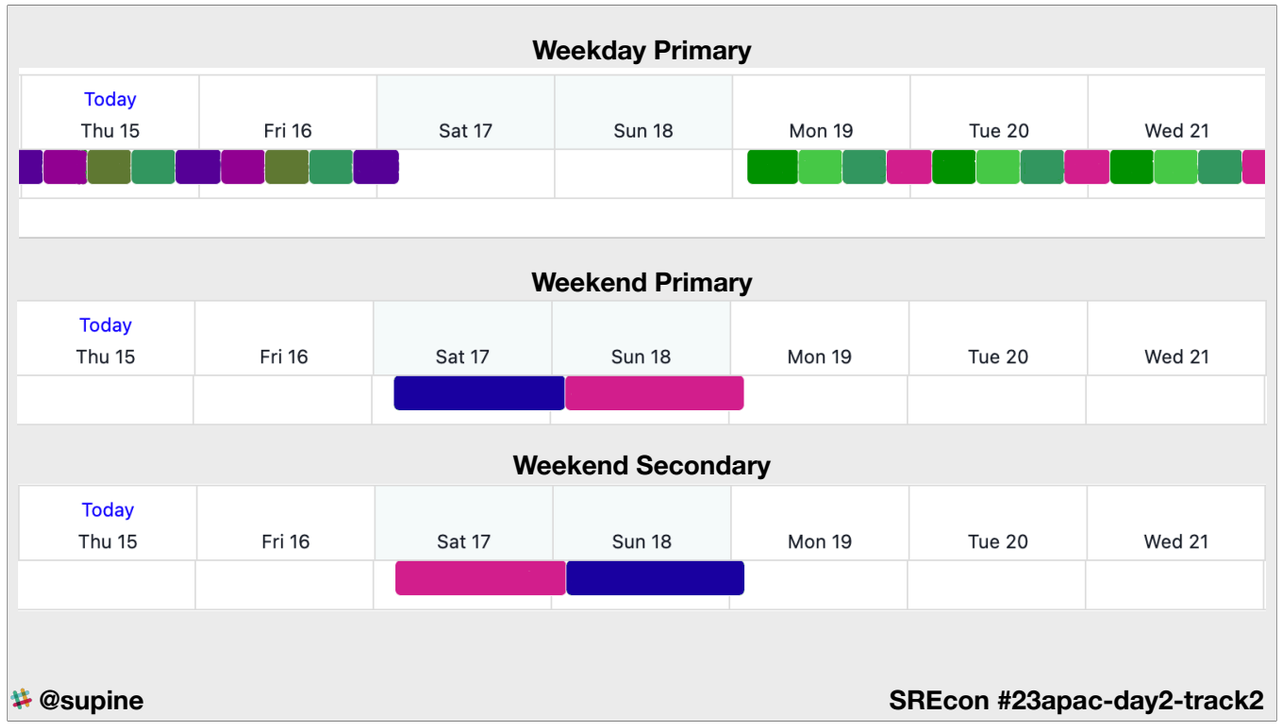
On a team I managed in the past we slowly grew until we had 12 people spread evenly across 4 different time regions.

That enabled us to do 6 hour shifts during the week when team members were already working. Everyone had primary shifts every third week.

But for the weekends we all agreed we'd prefer around-the-clock shifts spaced further apart.

Since I've changed teams they have adjusted it further, the primary and secondary weekend shifts swap halfway through the weekend.

But it means that you are on-call on the weekend only once in six weeks.



Not every on-call team needs this level of complexity.

It can be hard keep a long-term overview of when your on-call shifts actually are.

You might not have the alert load and workload that requires such frequent rotation.

You also can only work within the constraints you have.

It’s easy for me to stand here and say “grow your team, hire more people” but getting budget for that is not straightforward in good times and especially not in the economic and industry situation we have today.

Recovery Time

We’ve paid a lot of attention to when people will be on-call but it’s also important to spend some time on what happens after their shift ends.

Your on-call team members will need some recovery time.

Recovery Time

This is the time they need to recover from a lack of sleep.

This is the time they need to decompress after a stressful shift.

This is the time they need to catch up on their personal life that they put on hold during the shift.

Recovery Time

You need to consider both the short-term and long-term recovery.

Short-term thinking would be “can we expect this person to turn up to work today after being on-call overnight?”

Long-term recovery would be focused more on things like “when would it be reasonable for their next on-call shift to occur” all the way through to “this person probably needs a holiday sometime soon”.

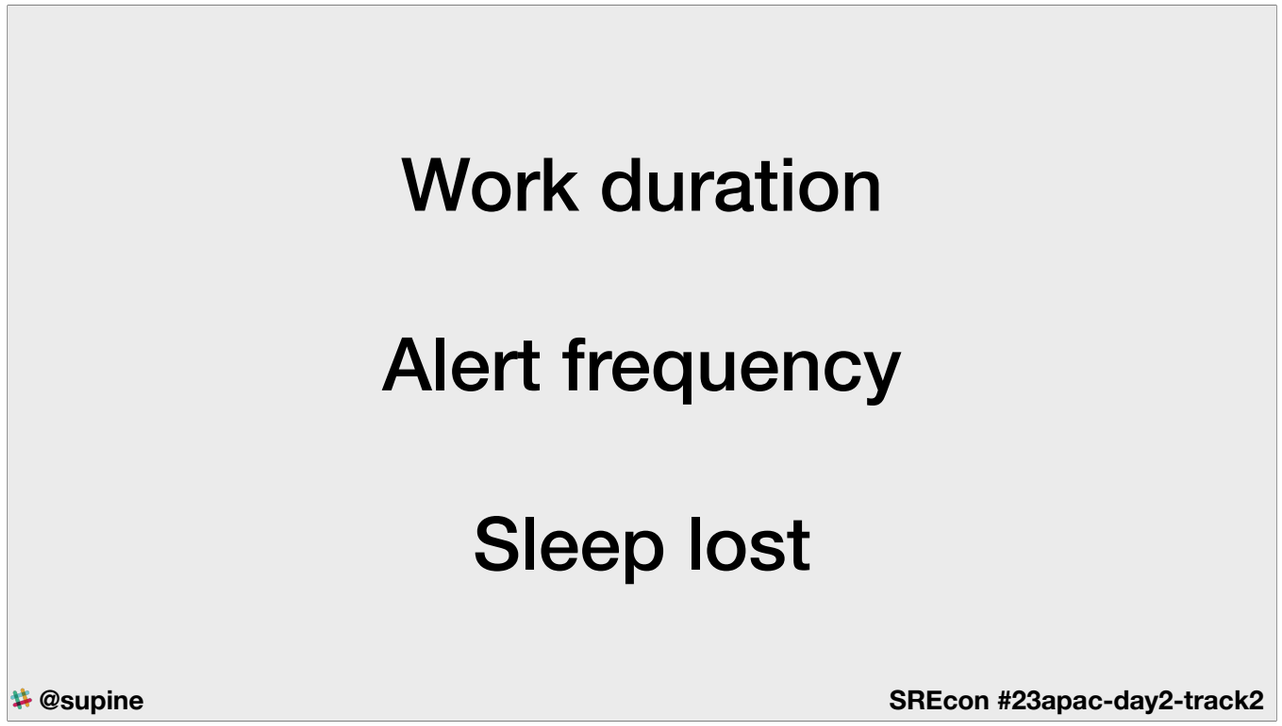
Circuit Breakers

We’ve talked a bit about improvements we can make to the on-call shift plan.

But we also need to talk about when fatigue management requires us to throw the plan out.

Sometimes the workload of an on-call shift is so bad that you need to relieve your primary on-call person, swapping in someone else either temporarily or for the entire remainder of the shift.

It’s important to think about and articulate what kind of signals are available at your company to help decide when this might be.



Here are some signals you could consider including in either a formal or informal policy.

During this on-call shift how long has the person been working, looking at both individual interventions and the cumulative time across all of them.

Even if the total amount of work time is low, someone might still have alert fatigue due to a high frequency of alerts that can be handled quickly.

And you can try to judge how much sleep they have lost, both in time spent working and in the delay in falling back asleep.

Realistic SLAs / SLOs

What other ways can be improve the on-call experience?

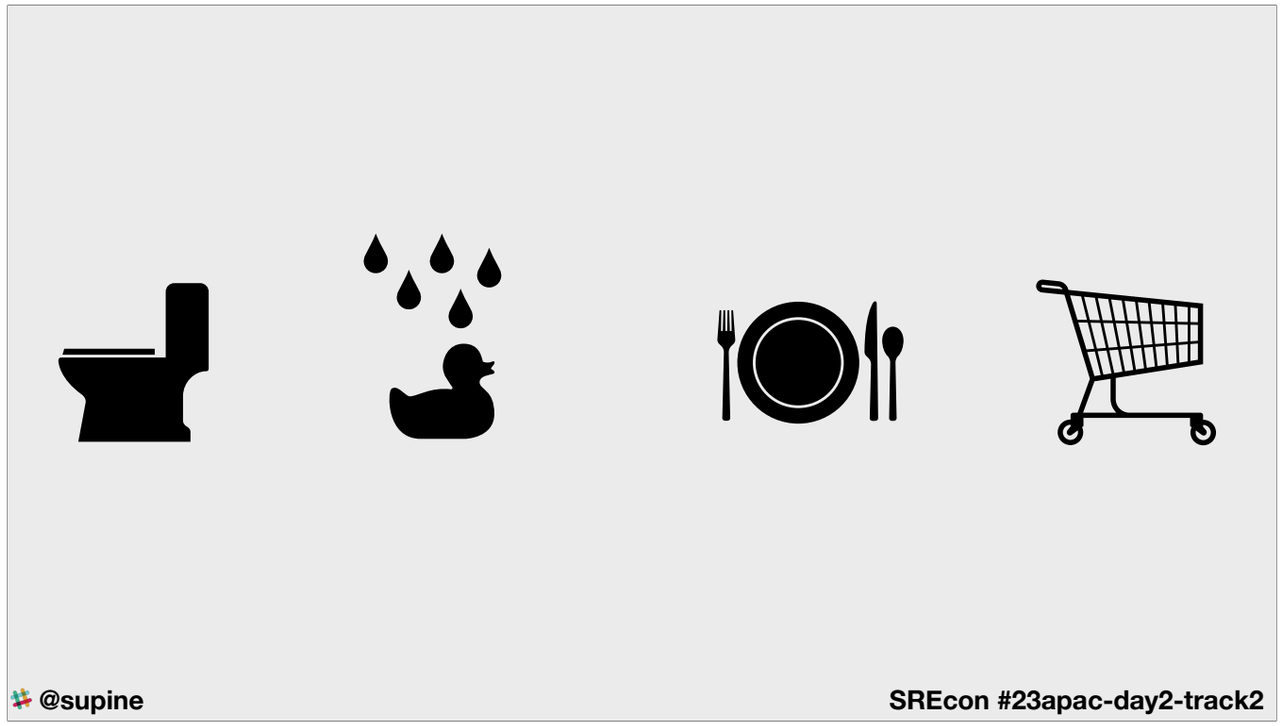
Companies will have certain expectations around how quickly someone on-call needs to respond once alerted.

The timings can cover various checkpoints like “time to acknowledge”, “time to laptop” and “time to start of work”.

What does your company think is a reasonable time between receiving an alert and starting to work on the problem?

30 minutes? 20? 10? 5?

To help judge how realistic an SLA is let us compare it to some things that might delay your response time.



An average visit to a toilet takes 3-4 minutes.

The average person spends 8 minutes in the shower.

Eating a proper meal should take you at least 20 minutes.

An average trip to a supermarket takes 40 minutes.

Some of these examples are easier to interrupt than others, but you can start to appreciate how aggressive SLAs can be incompatible with humans just living their life.

You almost certainly can’t leave the house without taking your laptop.

Actionable alerts

So now we have realistic expectations of when they will start working, but can they actually fix the problem?

Was the alert specific enough so it was routed to the appropriate team?

Could automation have attempted remediation before escalating it to a human?

Does the problem even need to be fixed immediately? Or can the degraded state be tolerated until waking hours or even business hours?

Timely and persistent remediation

Ideally alerts are not repetitive but if that does happen you need a clear path to timely and persistent remediation of the underlying problems.

This can take many forms, everything from simple bug fixes to finally decommissioning a legacy system.

But the remediation needs to be prioritised appropriately to stop the repetitive alerts impacting the on-call team.

Nothing drives someone to burnout like receiving yet another alert triggered by a problem that has persisted for months.

Business impact

One of the ways you can get help prioritising remediations is demonstrating the business impact caused by the problem.

Look for metrics that highlight the damage.

Loss of revenue.

Contracts at-risk or cancelled.

Increased costs.

Lost work hours.

Reputation loss.

Business impact

More generally, digging into business impacts can help you reframe how the rest of the company thinks about SRE and technical operations.

On-call is often simultaneously praised for how valuable it is while still being treated as a cost-centre where budgets can be held down or even reduced.

By highlighting the influence on-call has on those business impacts, you might be able to position the team as a protector of revenue and gross margin, worthy of investment.

On-call observability

You should also try to collect on-call specific metrics and logs.

When did alerts fire, who were they routed to, how long to acknowledgement, did the alert need to be resent, did it go to an escalation level.

Was this an automated alert from a monitoring system or a human paging an escalation person or subject matter expert.

Try to capture as much detail as possible without causing high cardinality problems.

On-call observability

A metric that is good to have, but is non-trivial to create, is “what was the local time for person who received the alert”.

This is the one time that UTC lets you down, you actually want to know if it was 3am in their timezone.

Compensation

The last topic I'd like to cover today is how your on-call team members are compensated for their work outside business hours.

A simple form of compensation is just paying them extra for it, some mix of a standby rate for making themselves available and a working rate for the time they actually spend on investigations and interventions.

This approach makes it clear how they are rewarded for their sacrifice and can easily handle shift swaps, particularly when they result in an imbalance in the on-call shifts across the team.

However this can create the wrong incentives, reducing balance and fairness.

Some folks might take on too much on-call because they want or need the extra money. Others might feel justified in reducing their on-call contributions because they are sacrificing the compensation.

Compensation

Some companies don’t explicitly pay extra for on-call, instead considering it included in the base salary.

Ideally the employee is aware of this when negotiating their salary and thinks that it is fair.

But even in this situation you probably still need a way to handle exceptions.

Not all shifts will be the same workload, so you need to consider what you can offer the person who had a really bad shift.

Depending on how bad it was, you might let them start their work day late or take the whole day off.

You could also offer than a flex day that they could take off when it suits them best.

Or more than one day if the shift was extremely bad.

Humane
On-call

To wrap up I’d like to remind everyone a key component in our systems are the humans who engage with it and we need to take care of them.

Work / life balance is so much more than closing your laptop at 5pm, especially when on-call is involved.

Thank you!

supine.com/srecon23apac

marty@supine.com

@supine

…and that’s all I have for you today.

I hope I’ve been able to explain a little about what it takes to care for the humans involved in on-call work and given you some things to think about as you return to your own work next week.

You can find my slides at that URL.

Thanks for listening and feel free to chat to me in the hallways or contact me online.