Power Laws and Heavy-Tail Distributions in Hyperscale Microservices Architectures

Abhishek Tiwari ⓘ

Recent analyses of Meta and Alibaba's production microservices architectures identified patterns of heavy-tail and power law distributions. These patterns manifest in service scale, and request patterns, providing a glimpse into the inherent characteristics of large-scale distributed systems.

For details review of Meta and Alibaba's microservices ecosystem please see [1] and [2].

Understanding Heavy-Tail Distributions

A heavy-tailed distribution emerges when extreme values appear with significantly higher probability than in a normal distribution. Power law represents a special type of heavy-tailed distribution where the probability of an event decreases proportionally to a power of its magnitude. In these distributions, the tail decays more slowly than exponential decay, resulting in a higher probability of extreme events.

In the context of microservices, heavy-tailed distributions mean that while most services exhibit "normal" behavior in terms of scale, complexity, or resource usage, there exists a significant number of outlier services with extreme characteristics. Power law distributions specifically indicate that the frequency of these extreme cases decreases in a mathematically predictable way according to a power function.

Service Complexity and Endpoint Distribution

Meta's analysis presents a clear power law in service complexity, measured by the number of endpoints exposed by each service ($\alpha = 2.23$, $R^2 = 0.99$). This means while most services are relatively simple with few endpoints, there exists a small but significant number of highly complex services. Their frontend service www, for example, exposes over 11,000 endpoints, demonstrating the extreme end of this distribution.
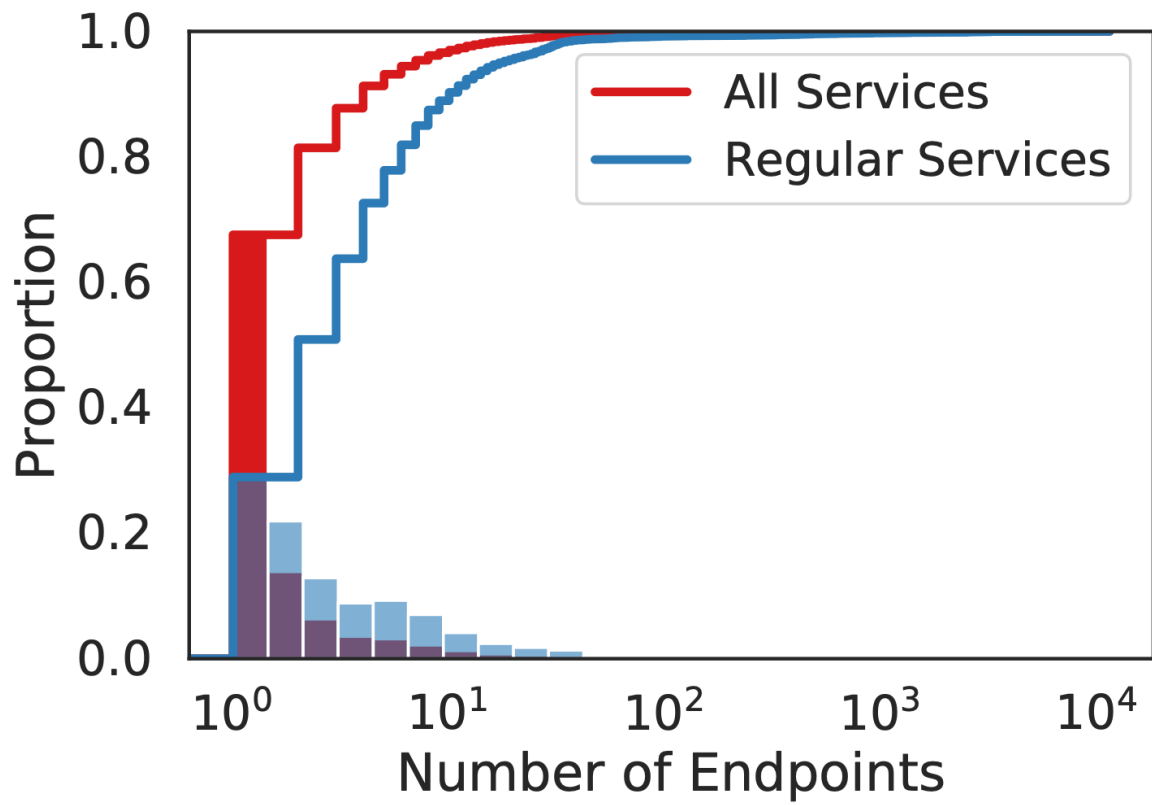
Figure 1: For Meta's microservices ecosystem, Service complexity measured by number of unique endpoints in a service shows a power-law distribution. Image credits Huye et. al

This pattern suggests a natural tendency for some services to accumulate functionality over time, possibly due to organisational dynamics (Conway's law) or technical constraints. It may also reflect conscious architectural decisions to consolidate related functionality within single services rather than further decomposition.

Call Graph Characteristics

Microservices call graphs have many characteristics (see [3]) and size of call graphs is one of them. Alibaba's analysis shows heavy-tailed distributions in call graph characteristics. While most service call graphs are modest in size, approximately 10% contain more than 40 unique microservices, with some graphs encompassing over 1,500 services.
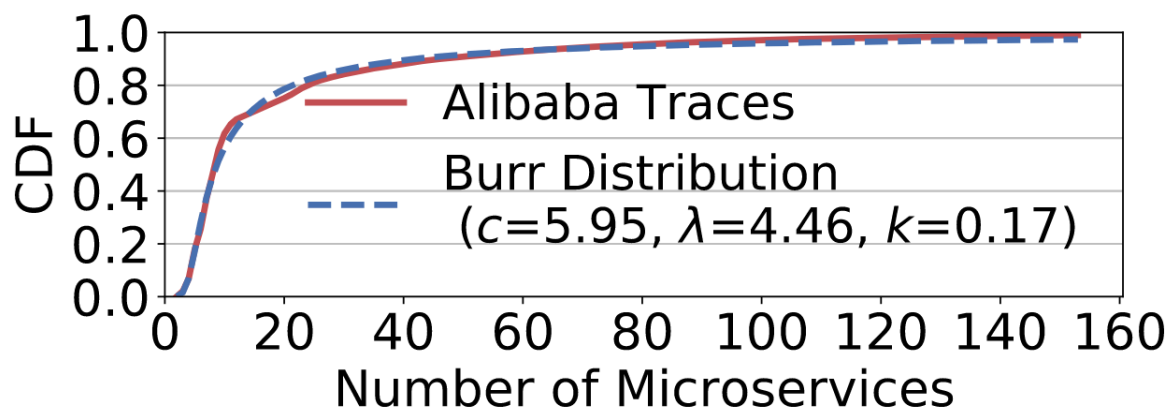
Figure 2: In Alibaba's microservices architecture, the number of microservices in a graph follows a Burr distribution. Image credits Luo et. al

Similarly, the number of calls within these graphs follows a heavy-tailed distribution or Burr distribution, with some graphs containing over 54,000 calls.
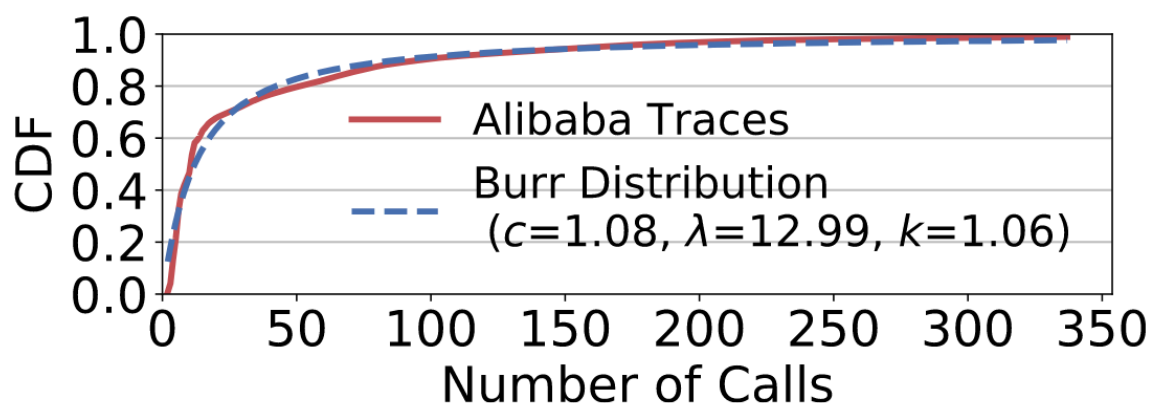

Figure 3: In Alibaba's microservices architecture, the number of calls in a graph also follows a Burr distribution. Image credits Luo et. al

These distributions have offer interesting learnings for system design and operation. Traditional capacity planning approaches that assume normal distributions may severely underestimate resource requirements for handling these extreme cases. Moreover, monitoring and debugging tools must scale effectively across several orders of magnitude of complexity.

Deviations

Interestingly, not all aspects of these architectures follow expected power law patterns. Meta's analysis found that despite there being a long tail of more complex services, their overall service dependency graph does not exhibit the power law relationships typically seen in large-scale networks (R^2 = 0.62). Similarly, the distribution of service instance counts shows no clear power law pattern (R^2 = 0.25).

These departures from expected behaviours suggest that architectural decisions and operational practices can sometime override "natural" network growth trends.

Conclusion

The heavy-tail and power law distributions in large-scale microservices architectures highlights fundamental patterns in distributed systems. These distributions naturally emerge from the interconnected nature of microservices and the varying complexity of business operations they support. An understanding of these distribution offers a simple mental model for microservices system design and resource allocation.

References

[1]   A. Tiwari, "Microservice Architecture of Alibaba," 2024, *Abhishek Tiwari*. doi: 10.59350/4fttv-pt832.

[2]   A. Tiwari, "Microservice architecture of Meta," 2024, *Abhishek Tiwari*. doi: 10.59350/7x9hc-t2q45.

[3]   A. Tiwari, "Unveiling Graph Structures in Microservices: Service Dependency Graph, Call Graph, and Causal Graph," 2024, *Abhishek Tiwari*. doi: 10.59350/hkjz0-7fb09.