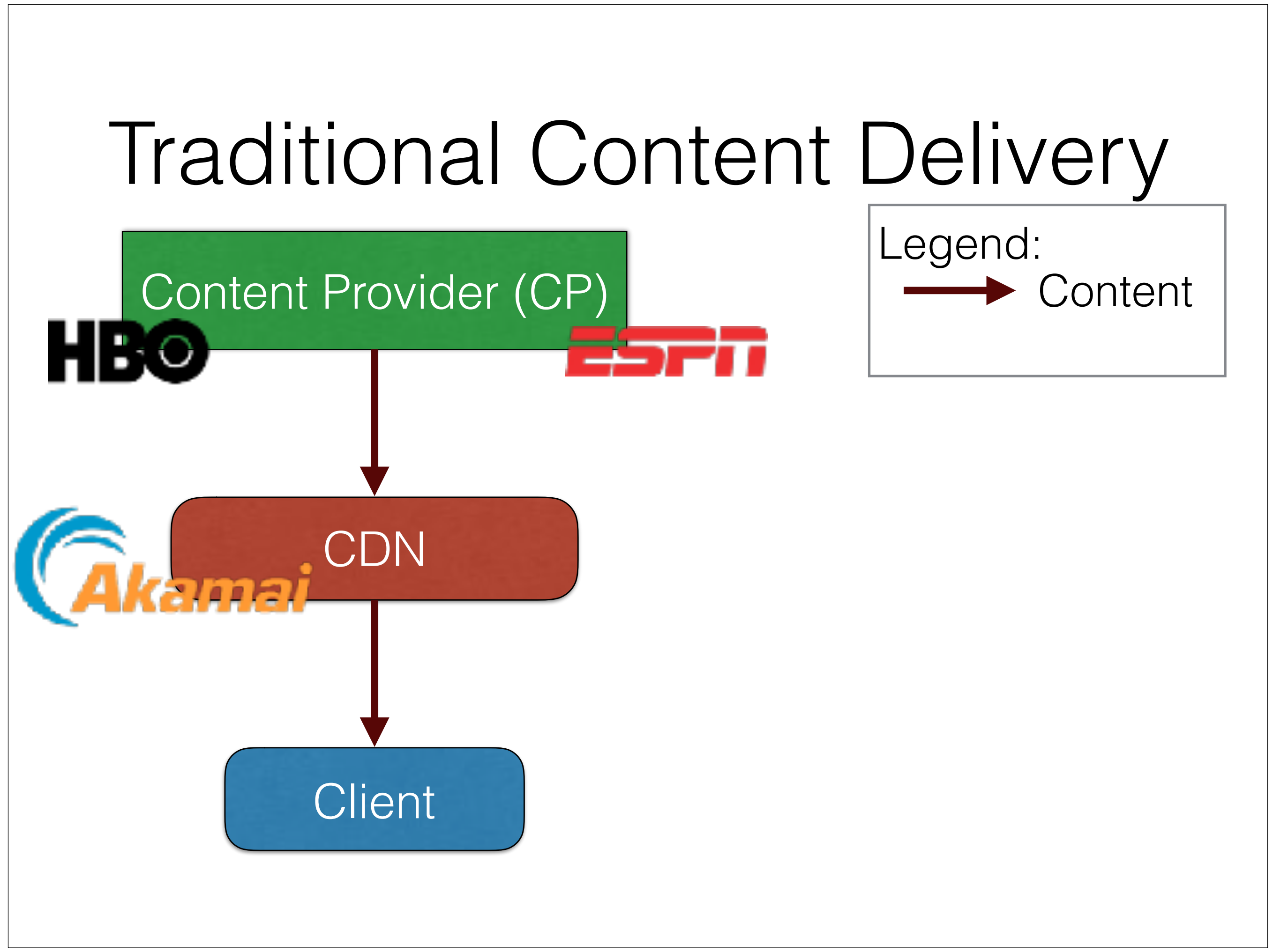
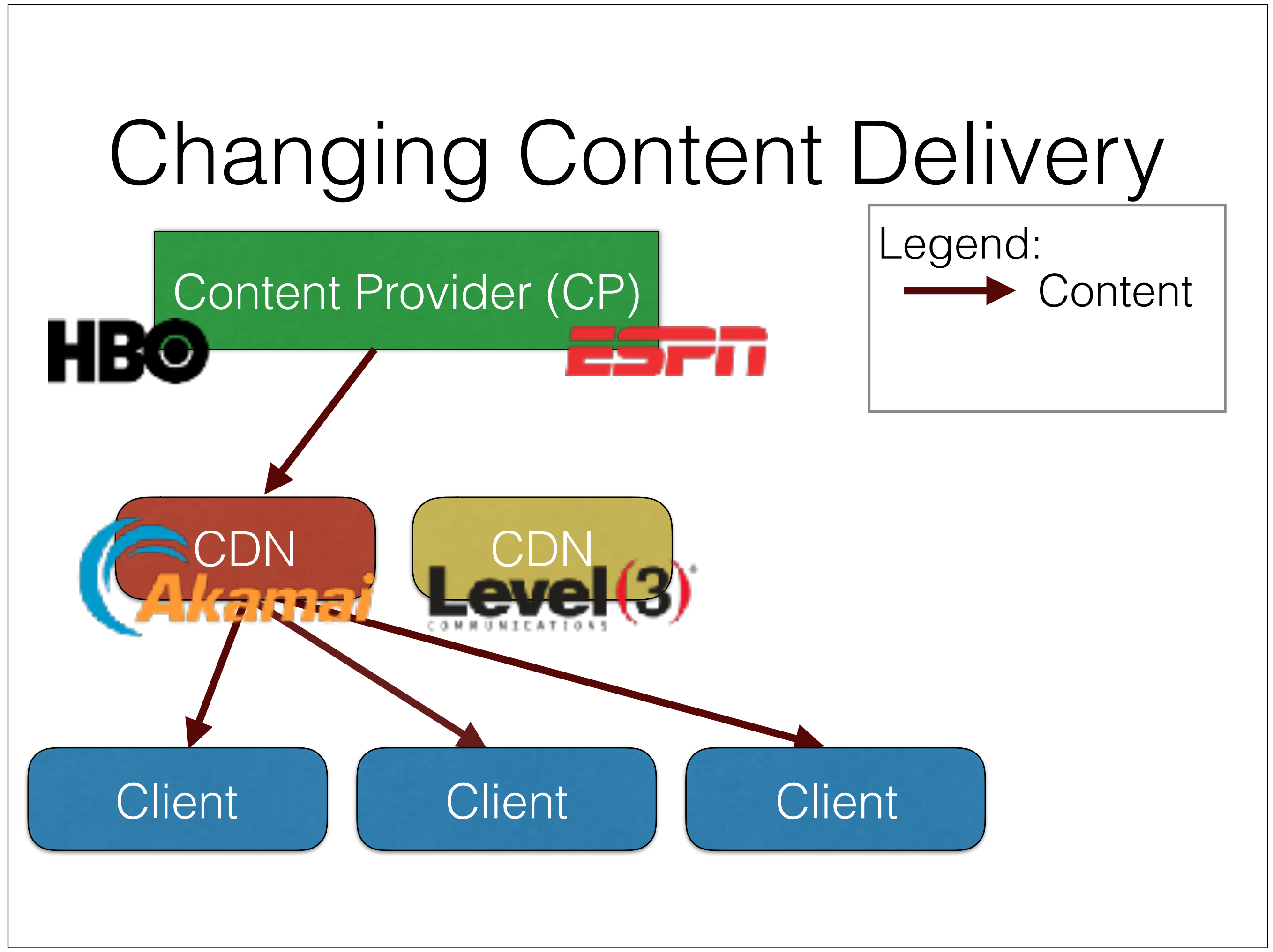
The Impact of Brokers on the Future of Content Delivery

Matthew K. Mukerjee, I. Nadi Bozkurt, Bruce Maggs,
Srinivasan Seshan, Hui Zhang

Hi, my name is Matt Mukerjee and I'll be presenting our work on "The Impact of Brokers on the Future of Content Delivery."

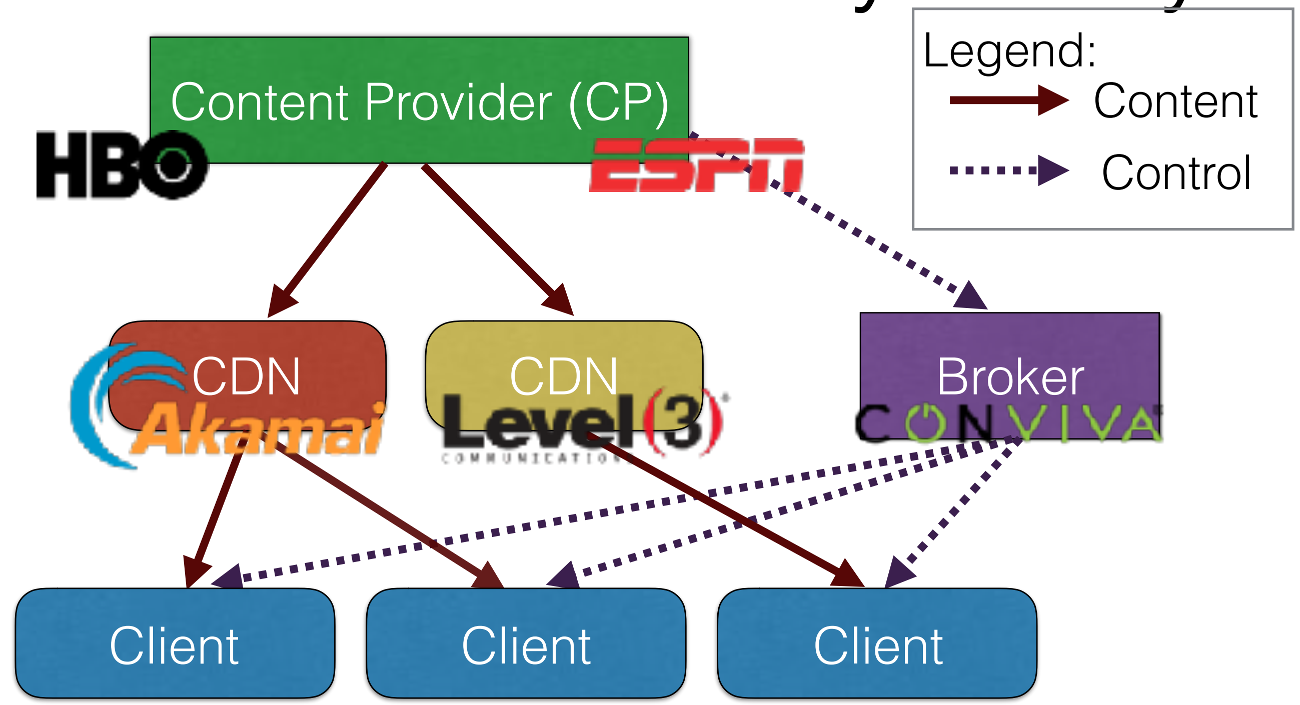


Traditional content delivery involves content providers (like ** HBO and ** ESPN), sending their content to CDNs (like ** Akamai), which ultimately deliver the data to clients. The picture is complicated by…

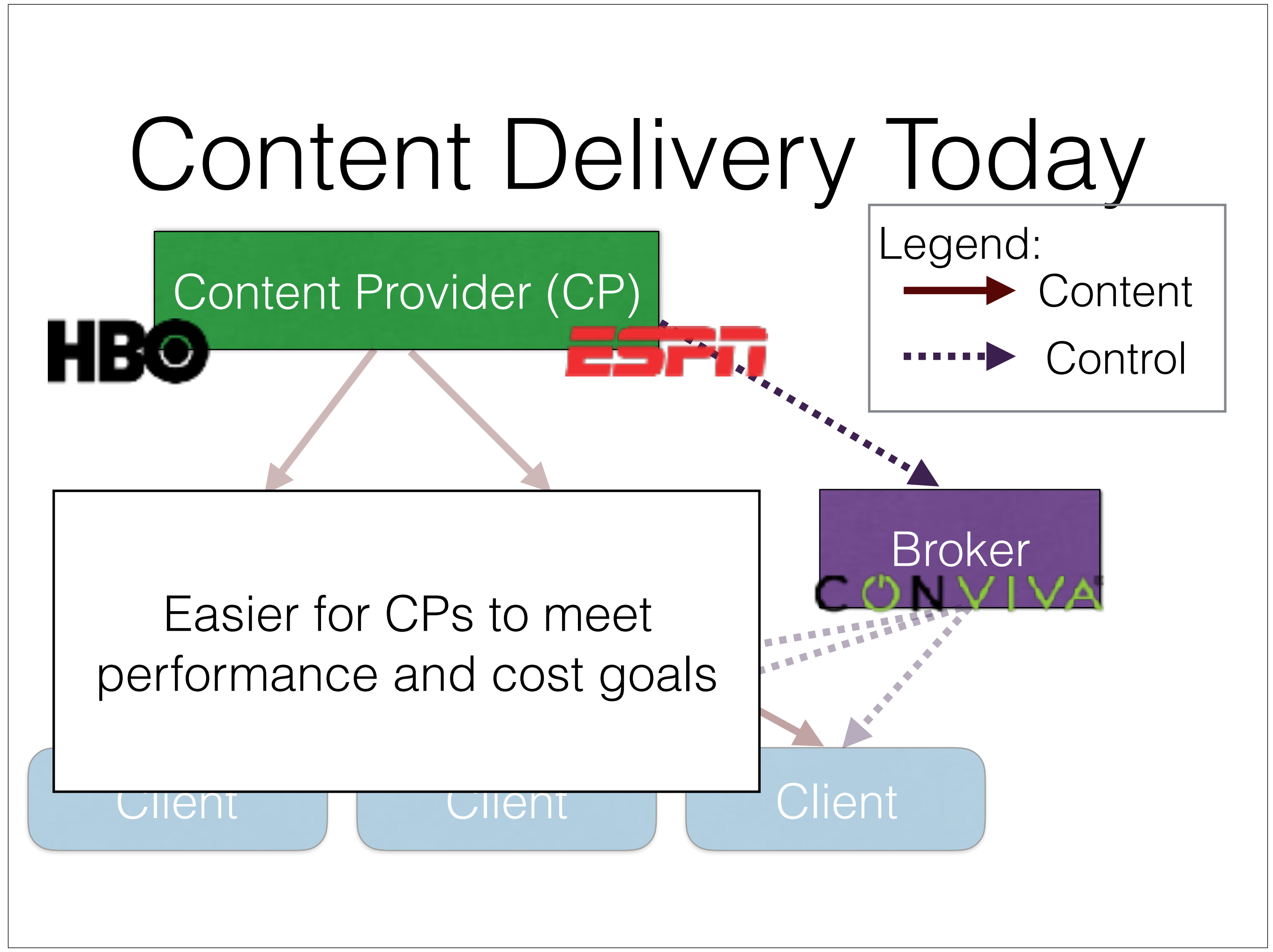


… many clients as well as ** other CDNs **. In order to make better use of the opportunities offered by stitching together multiple CDNs, an additional entity is involved in content delivery today, …

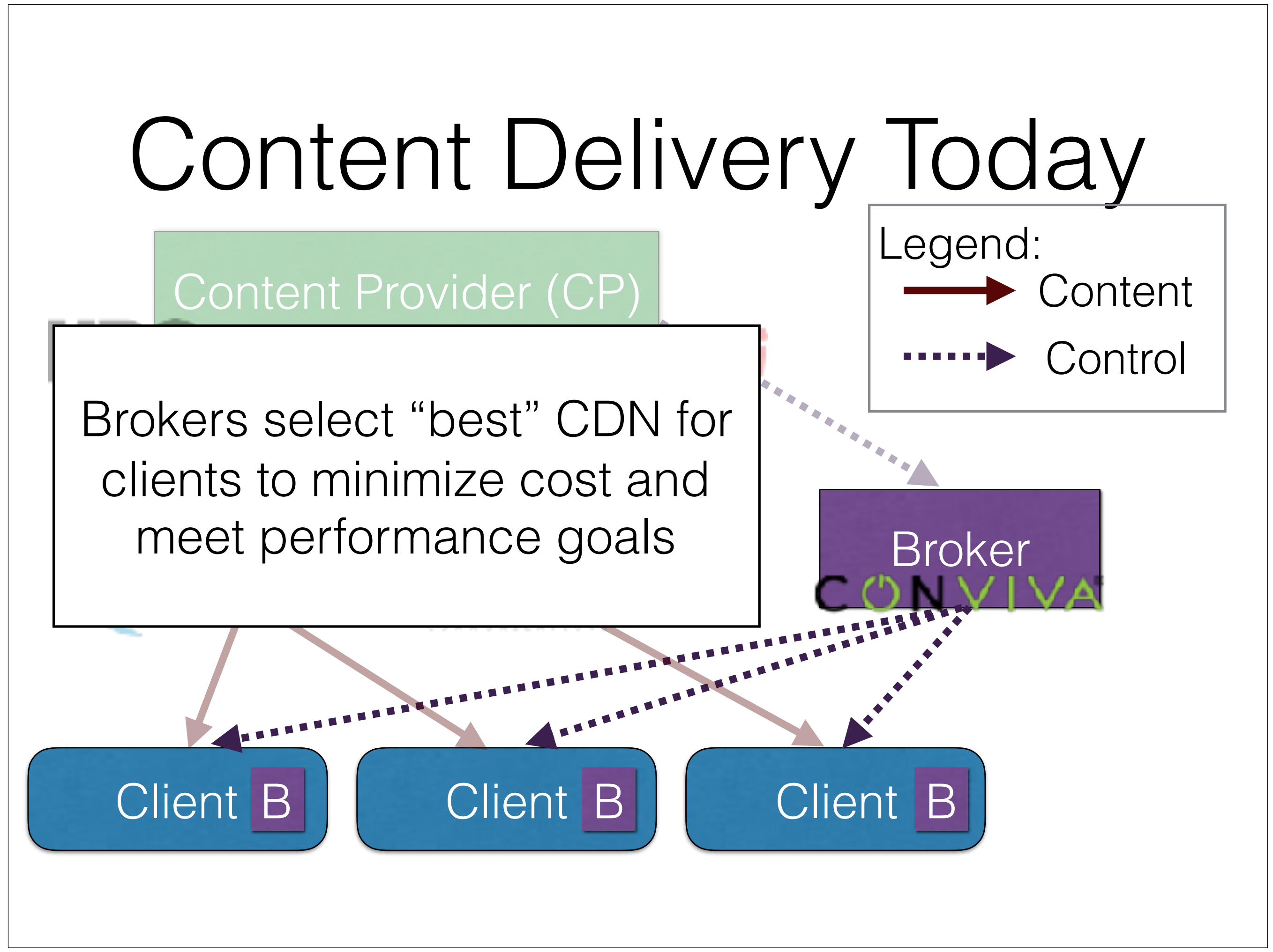
Content Delivery Today



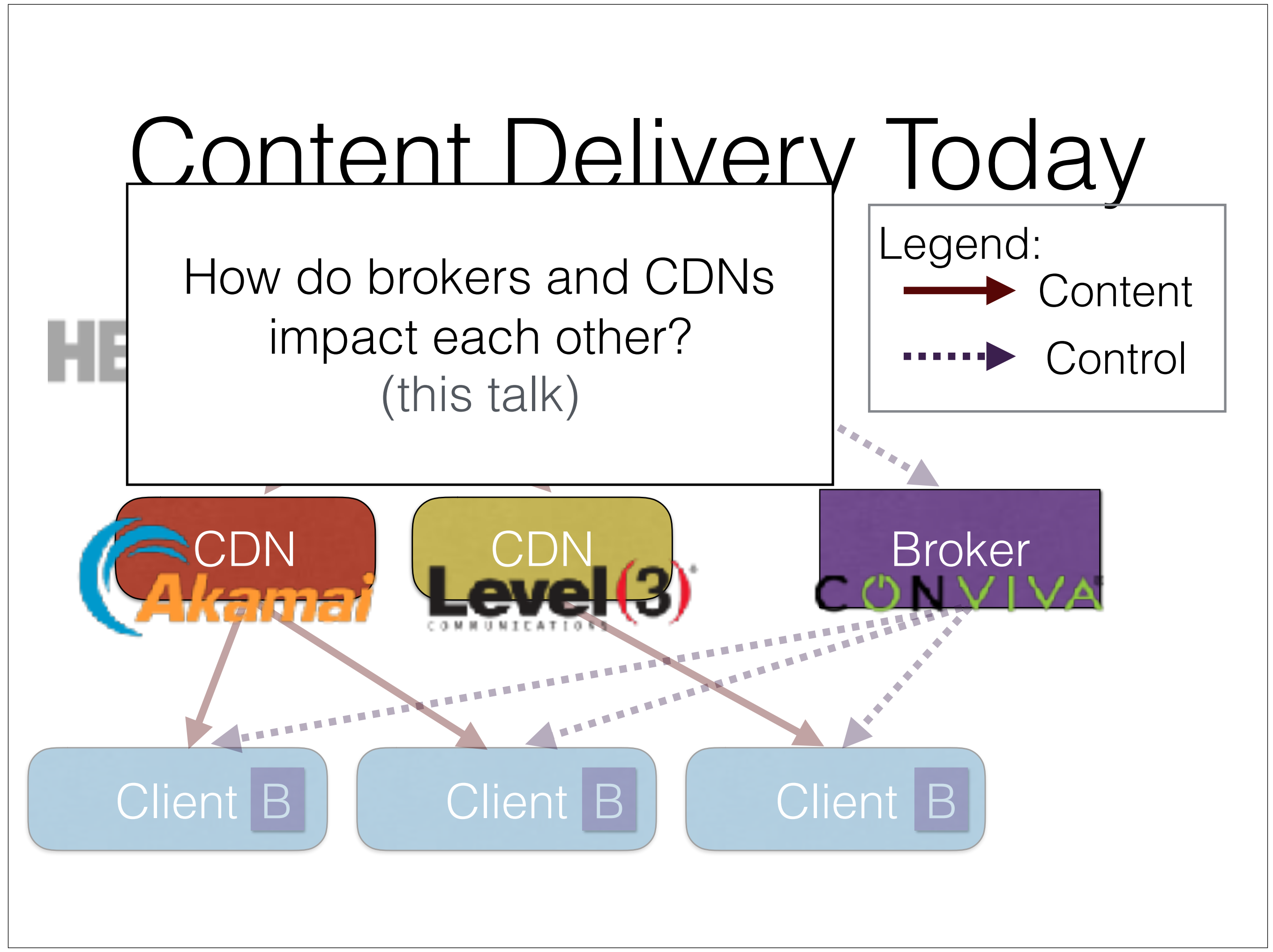
… called a broker (** the best known broker being Conviva). Brokers are purely a control plane entity that stitch together CDNs, …



… making it easier for content providers to meet performance and cost goals.



They do so by selecting the appropriate CDN for clients. Brokers run software on the clients (e.g., a video player on ESPN’s website) that contact the broker periodically to select the “best” CDN for the client based on things like device type, geographic location, and ISP. The “best” CDN may change over time.



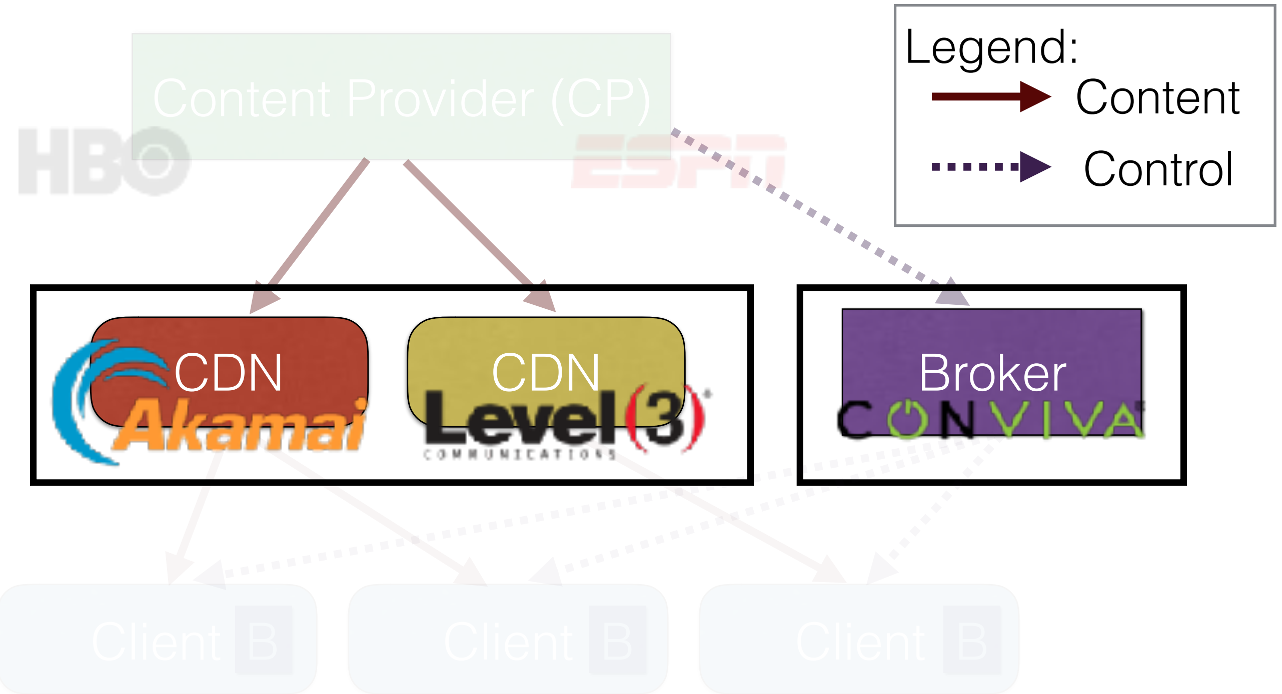
What we don't understand well is how the decisions made by the broker affect the decisions made by the CDNs and vice-versa. To exacerbate this— currently brokers and CDNs don't have an interface; they don't explicitly communicate with each other to make decisions, potentially leading to problems.

Contributions

- Identify challenges that brokers and CDNs create for each other by analyzing data from both
- Propose a CDN-broker interface based on an *ad exchange* that benefits both

In this work, ** we identify these problems by analyzing data from both, as well as ** propose an initial CDN-broker interface to fix these problems based on an ad exchange.

Potential Problems



First— potential problems: we group potential problems into two categories: ** problems faced by CDNs and ** problems faced by brokers. (Let's dig into these)

Potential Problems

CDN

- Brokers cause CDN traffic to be unpredictable at **short** and **long** timescales making provisioning difficult
- When traffic is unpredictable, flat pricing model makes profits unpredictable

Broker

- Coarse CDN-level selection complicate meeting CP goals
- Incomplete measurements complicate meeting CP goals
- Difficult debugging

CDNs face traffic unpredictability at both short and long timescales, making provisioning difficult. ** When traffic is unpredictable, CDNs flat pricing model makes profits unpredictable. Brokers face a different set of problems **, but in this talk, we're going to only…

Potential Problems

CDN

- Brokers cause CDN traffic to be unpredictable at **short** and **long** timescales making provisioning difficult
- When traffic is unpredictable, flat pricing model makes profits unpredictable

Broker

- Coarse CDN-level selection complicate meeting CP goals
- Incomplete measurements complicate meeting CP goals
- Difficult debugging

See Paper

… focus on problems CDNs face. For insight into problems faced by brokers, read through our paper. First, ** let's look at how unpredictable traffic makes provisioning difficult for CDNs.

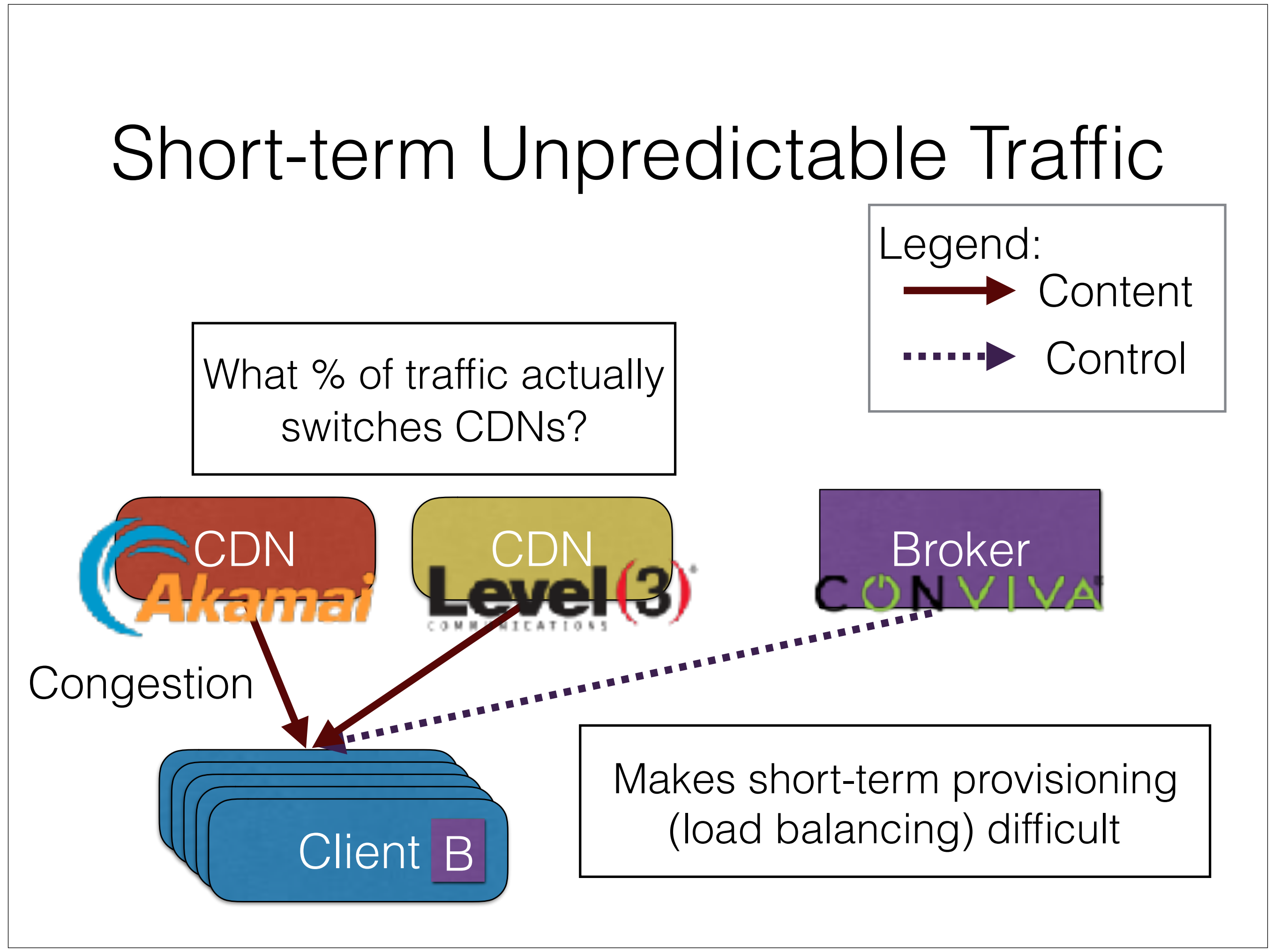
Potential Problems

- Brokers cause CDN traffic to be unpredictable at **short** and long timescales making provisioning difficult
- When traffic is unpredictable, flat pricing model makes profits unpredictable

- Coarse CDN-level selection complicate meeting CP goals
- Incomplete measurements complicate meeting CP goals
- Difficult debugging

See Paper

Specifically, let’s first focus on short term unpredictability, then talk about long term unpredictability.



Let's look at this example. Here we see a client getting content from Akamai, but ** now there's congestion. A broker can jump in (mid-session) ** and move this client ** to another CDN. Now imagine, instead of a single client ** this happens to a large number of clients. Clearly, moving large numbers of clients from one CDN to another ** makes short-term provisioning (i.e., load balancing) difficult for both CDNs. Does this problem actually happen in the wild though? Let's look at data from a broker to find out ** what % of traffic actually switches CDNs.

Short-term Unpredictable Traffic

40%

of video delivery sessions switched CDNs during lifetime

Makes short-term provisioning (load balancing) difficult

We got data from a large broker involved in video delivery. The data contains video sessions from clients over one hour. We find that ** 40% of sessions switched CDNs during their lifetime. There's a nice graph of this in the paper in detail. Thus when a broker is involved, ** CDN load balancing is potentially more difficult.

Potential Problems

CDN

- Brokers cause CDN traffic to be unpredictable at **short** and long timescales provisioning difficult
- When traffic is unpredictable, flat pricing model makes profits unpredictable

Broker

- Coarse CDN-level selection complicate meeting CP goals
- Incomplete measurements complicate meeting CP goals
- Difficult debugging

See Paper

Now that we've seen how short term traffic unpredictability affects provisioning, …

Potential Problems

CDN

- Brokers cause CDN traffic to be unpredictable at short and **long** timescales provisioning difficult
- When traffic is unpredictable, flat pricing model makes profits unpredictable

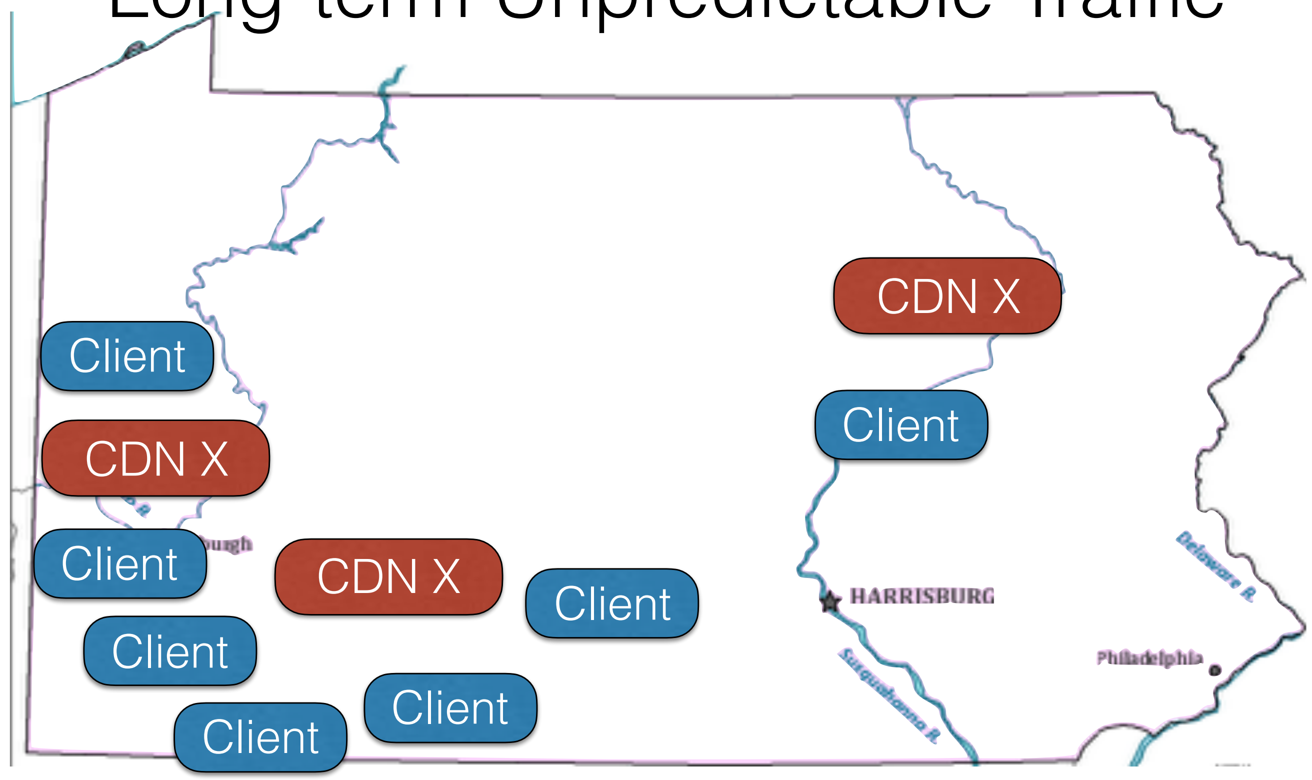
Broker

- Coarse CDN-level selection complicate meeting CP goals
- Incomplete measurements complicate meeting CP goals
- Difficult debugging

See Paper

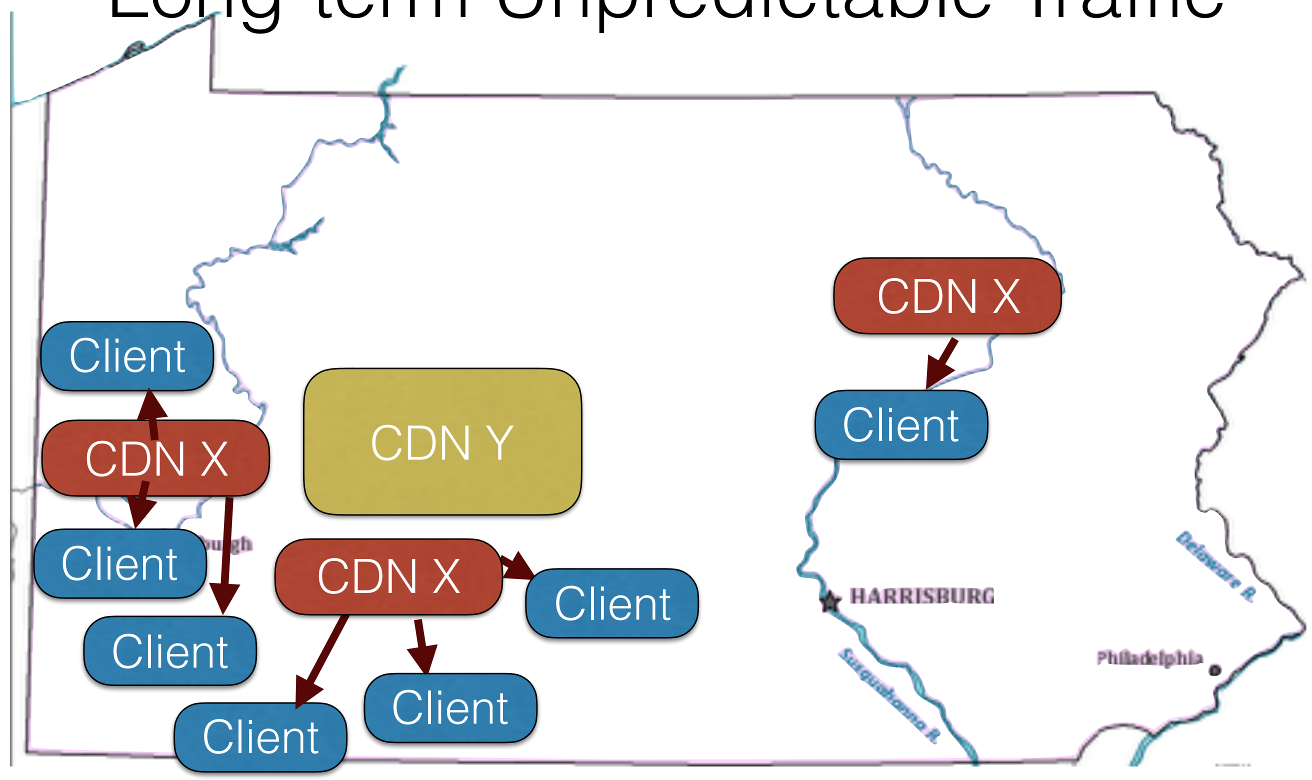
… let's look at how long term unpredictability.

Long-term Unpredictable Traffic

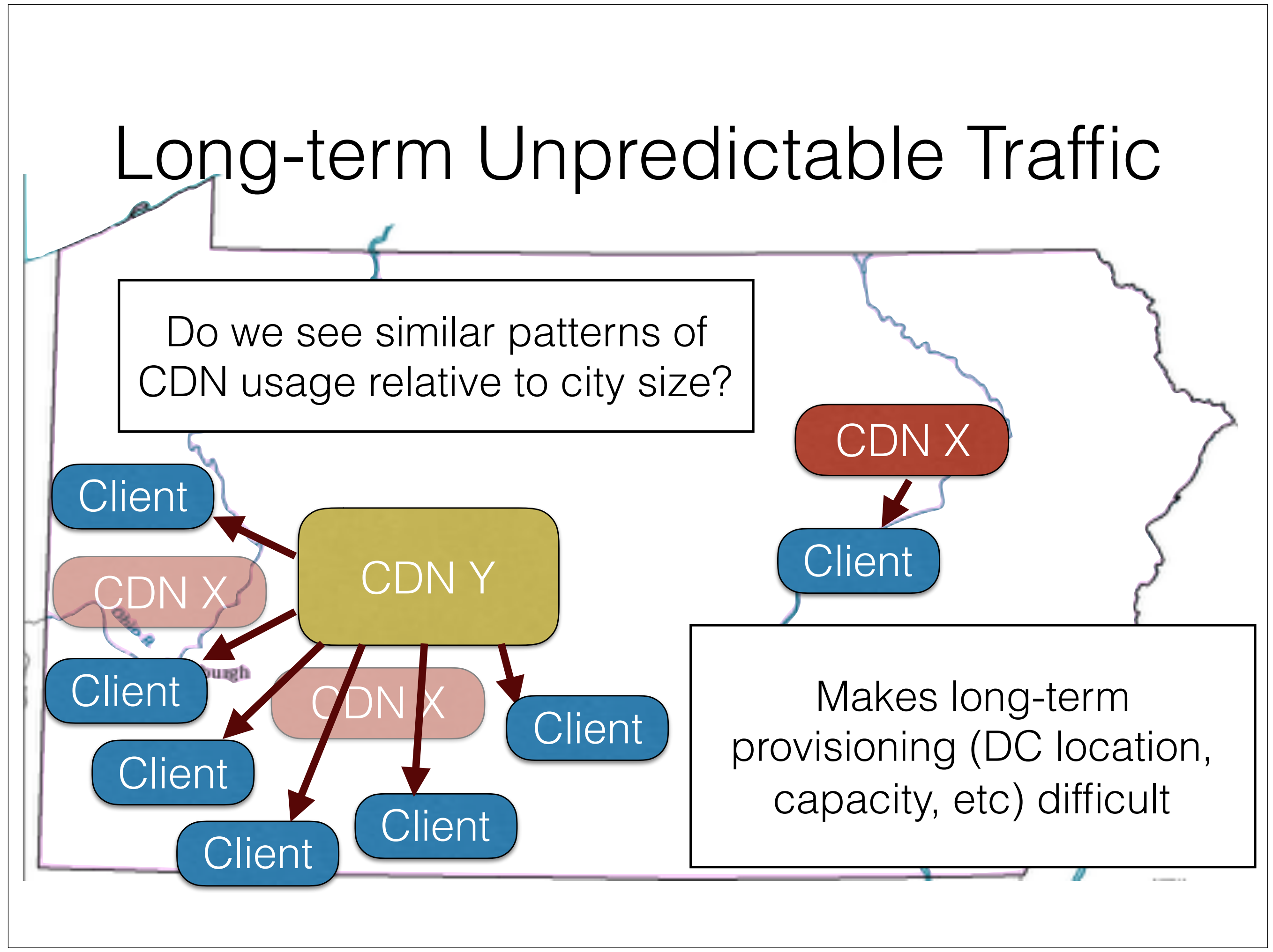


Let's step through another hypothetical example. ** Here we see many clients in Pittsburgh, and ** one client in this rural area. ** ** Here we see CDN X's clusters.

Long-term Unpredictable Traffic

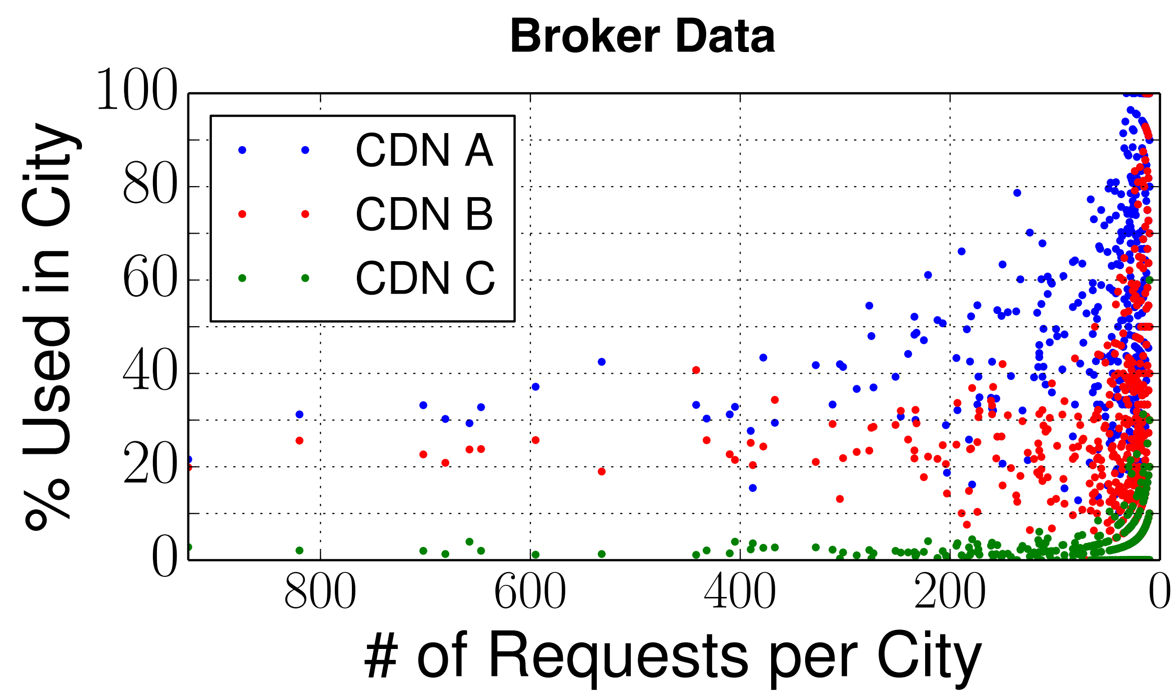


CDN X builds many delivery clusters so that their clusters are always close to clients, providing good performance. ** CDN Y takes an alternate approach, opting for fewer, high-capacity clusters with a cheaper price.

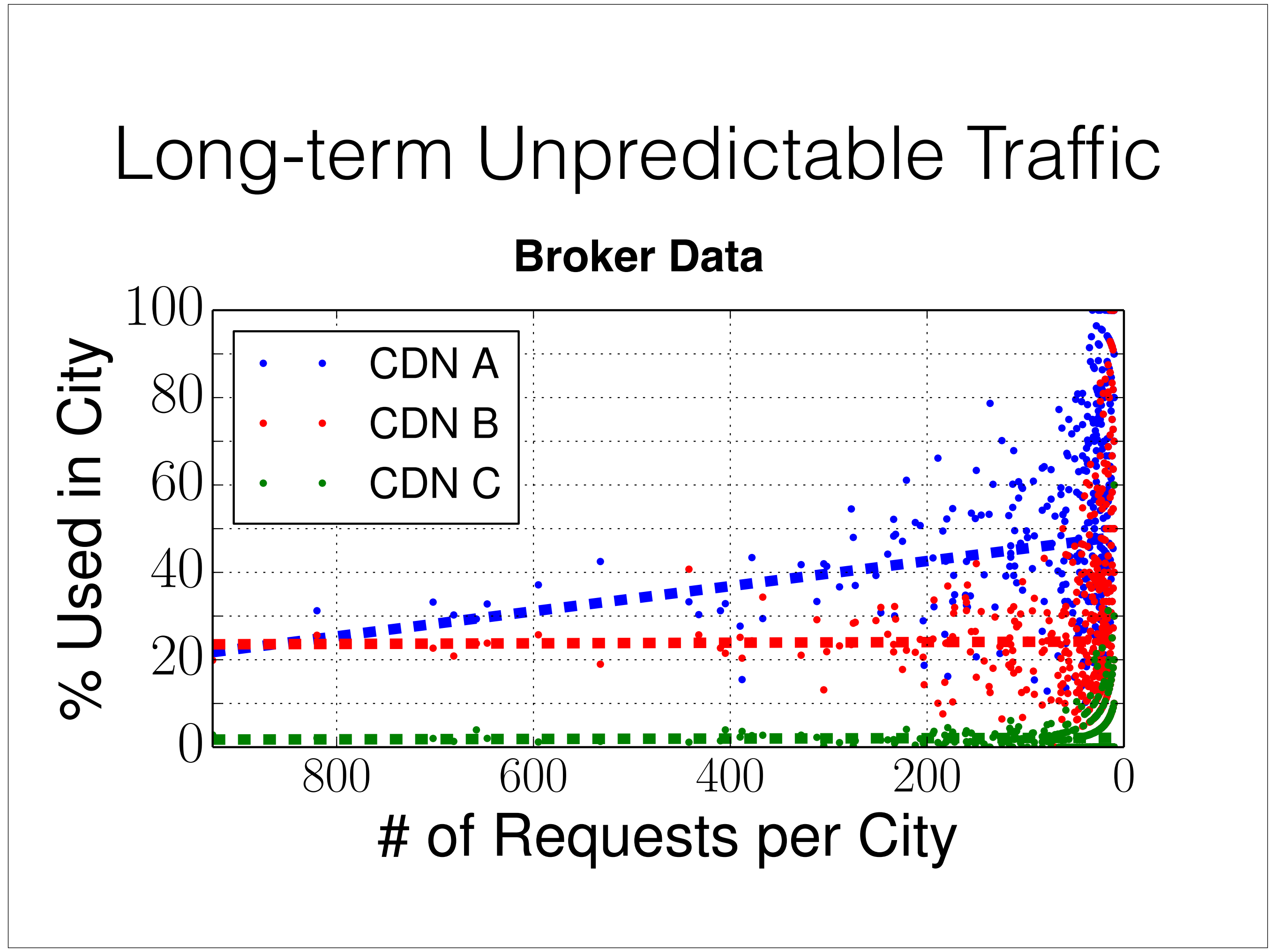


A broker sees that CDN Y can provide adequate performance at lower price, moving all the clients in the Pittsburgh area to CDN Y's cluster. In effect, the broker pushed CDN X out of the major city, only using it in rural areas. This goes against traditional provisioning wisdom— there is no longer positive correlation between number of clients in a region and the number of delivery clusters that should be placed in that region, ** in effect making long-term provisioning difficult (e.g., datacenter location, capacity planning, etc.). To see if this is an issue in practice, let's look at broker data ** to look for similar patterns in CDN usage relative to city size.

Long-term Unpredictable Traffic

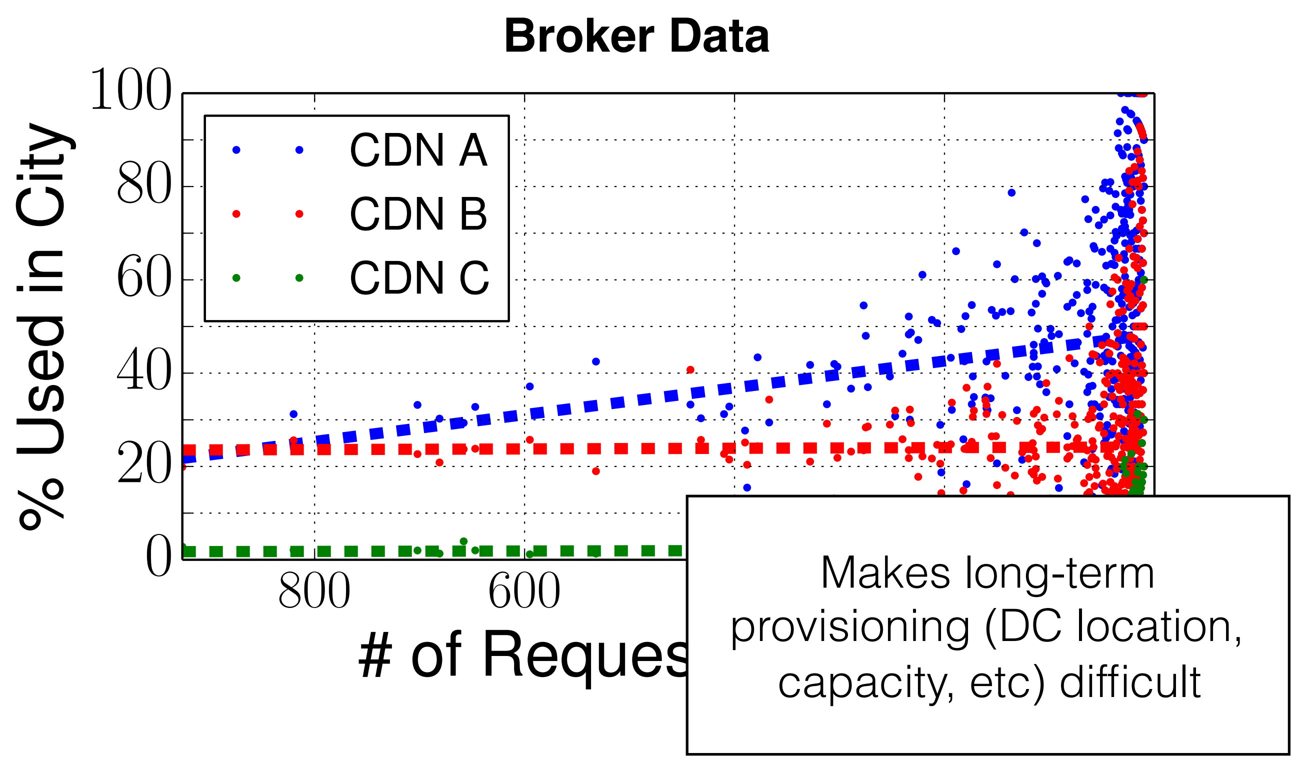


On the x-axis, we see cities sorted from large on the left to small on the right. On the y-axis we show which CDNs served clients in those cities as a percentage. The color series show the three CDNs explicitly labeled in our data as A, B, and C. The rest of the clients were served by “Other CDNs” which were grouped together in the data and are not plotted. To better understand the trends…



… we plot best-fit lines over the data. The key takeaway is that CDN A is being used pushed towards specialized “small city” delivery. This CDN is similar to “CDN X” in our previous example: this CDN builds many delivery clusters both in large and small cities, but is more expensive when compared to its competitors. Thus, when other CDNs can provide adequate performance (in big cities— on the left), the more expensive CDN A is avoided. But in smaller cities (on the right), the performance gain of having a cluster closer outweighs the increase in price, thus CDN A is used more.

Long-term Unpredictable Traffic



As mentioned, this make long-term provisioning difficult, as client location is no longer a good indicator for proper datacenter placement.

Potential Problems

CDN

Akamai

- Brokers cause CDN traffic to be unpredictable at short and long timescales making provisioning difficult
- When traffic is unpredictable, flat pricing model makes profits unpredictable

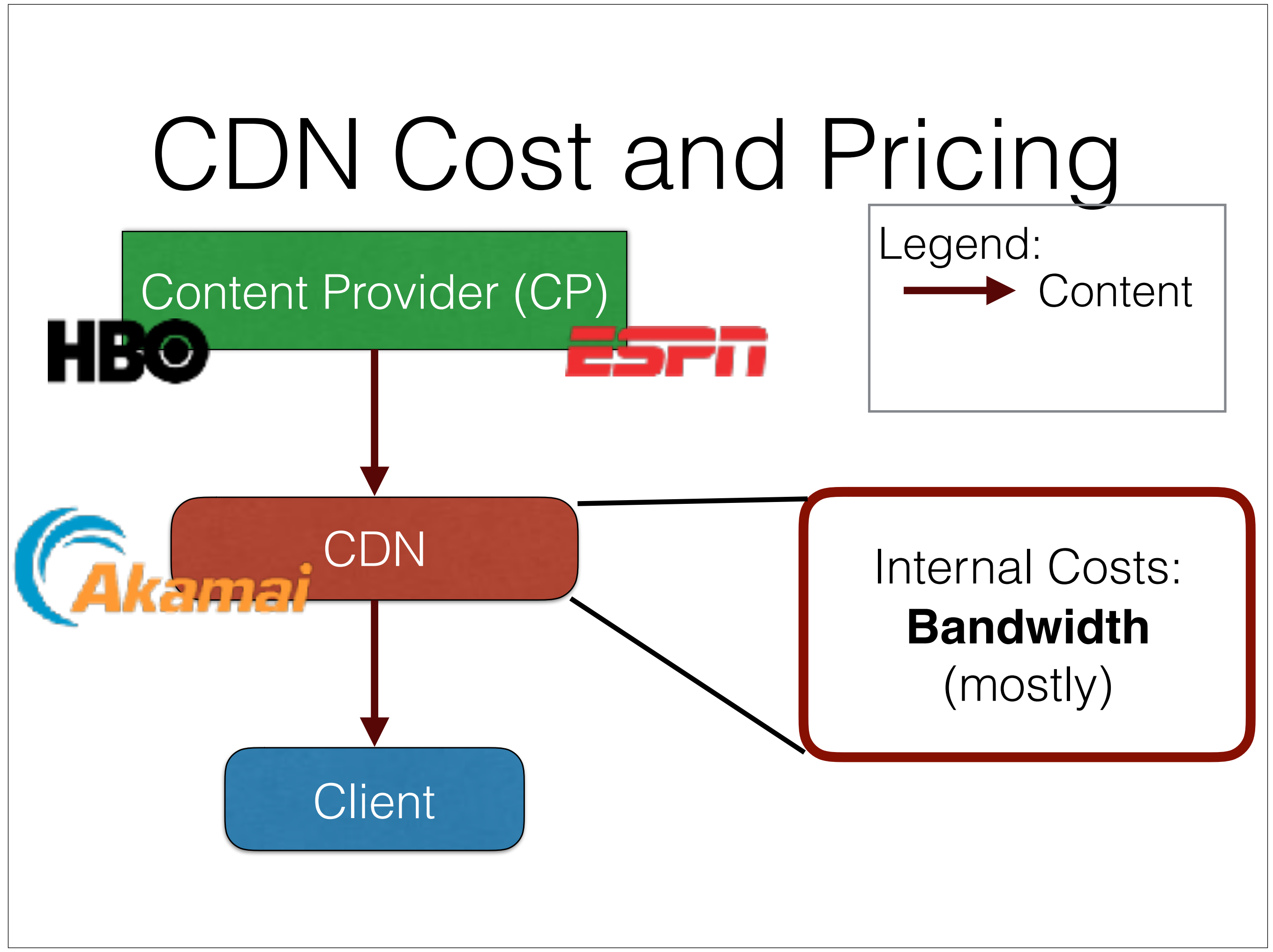
Broker

CONVIVA

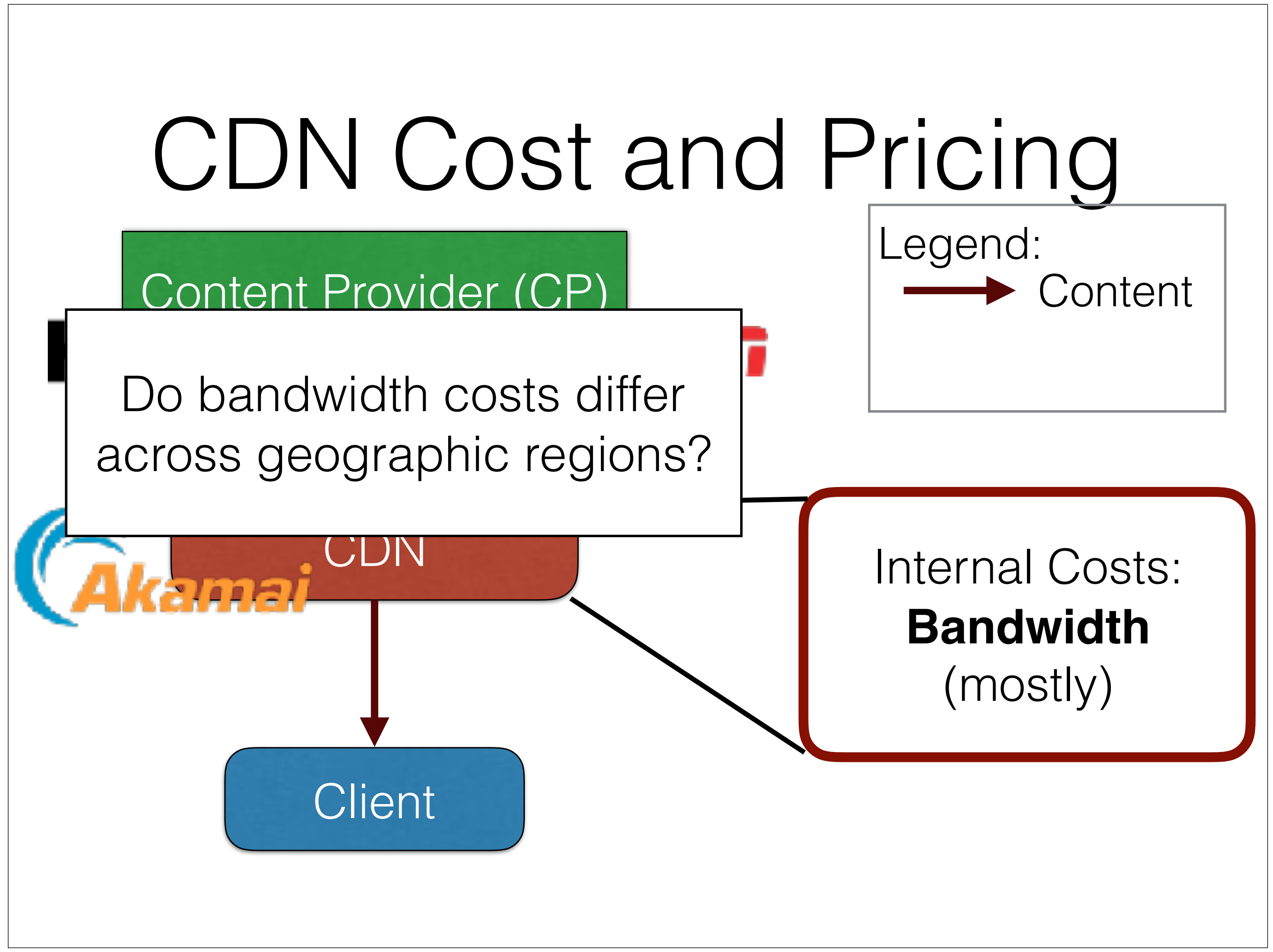
- Coarse CDN-level selection complicate meeting CP goals
- Incomplete measurements complicate meeting CP goals
- Difficult debugging

See Paper

Now that we understand why traffic has become unpredictable, let’s understand why this impacts CDN profits.



To understand CDN profits, we need to understand their internal cost and revenue. ** We were told by a large CDN that their internal cost comes predominantly from paying ISPs for bandwidth.



A natural question is if bandwidth costs differ across geographic regions.

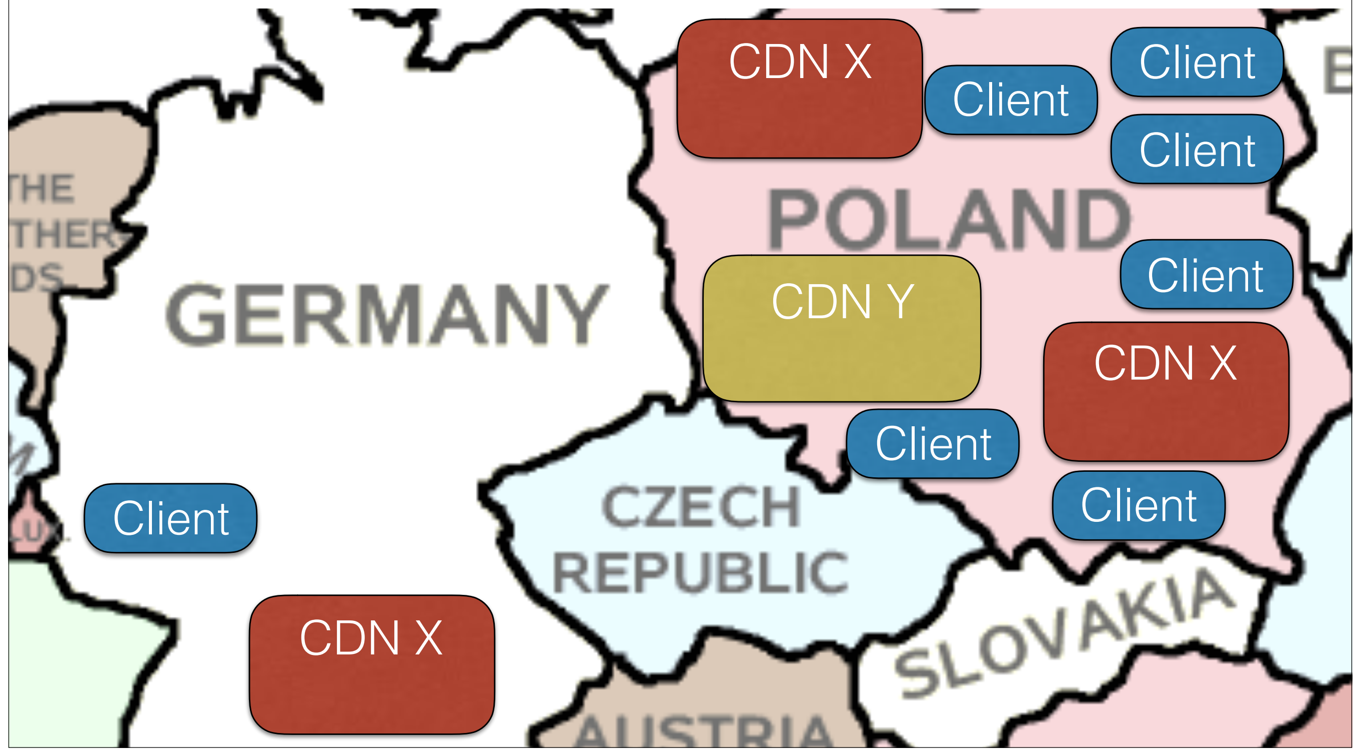
CDN Cost / Byte Delivered

30x

difference in cost per byte
between the most expensive
and least expensive countries

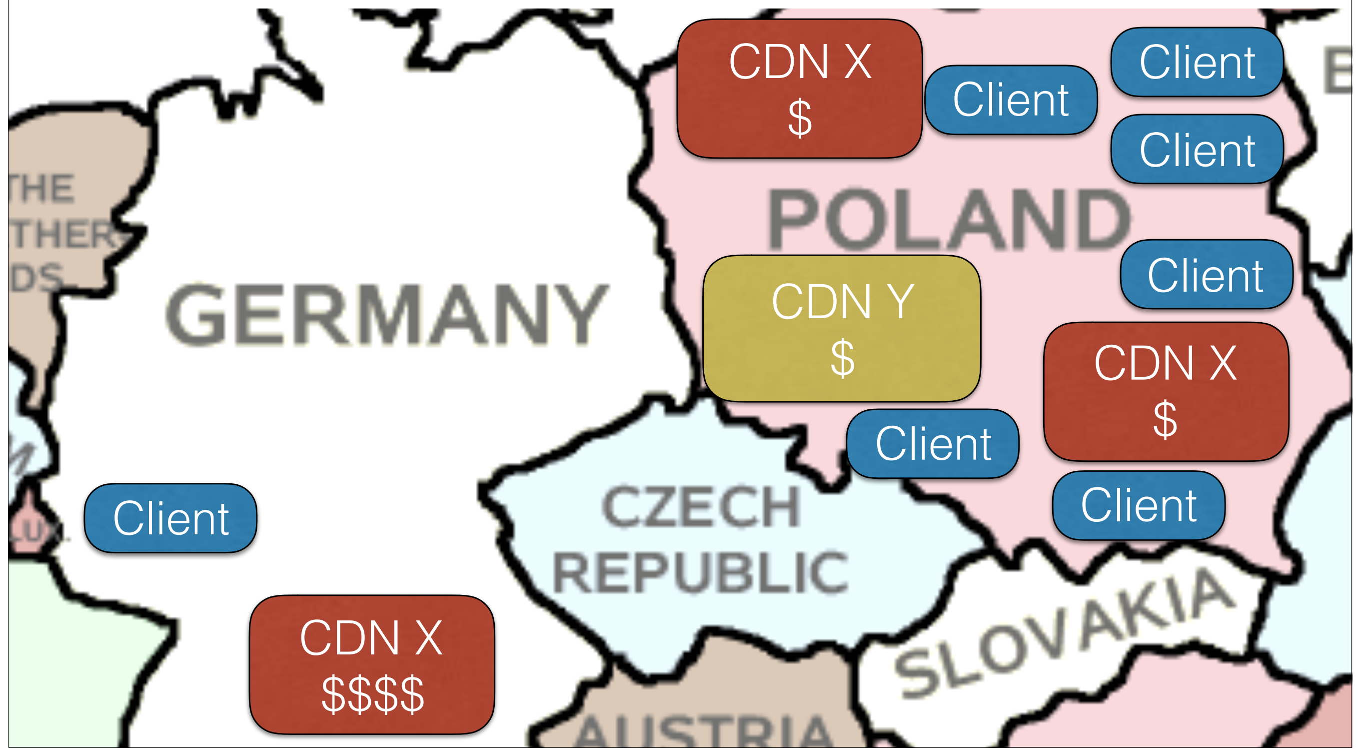
We got data on the cost per byte delivered from a major CDN for the top 20 countries with the most requests. There was a ** 30 times difference in cost between the most expensive and least expensive country. I want to explain how we're going to represent these internal costs, so let's go through another hypothetical example.

CDN Internal Cost

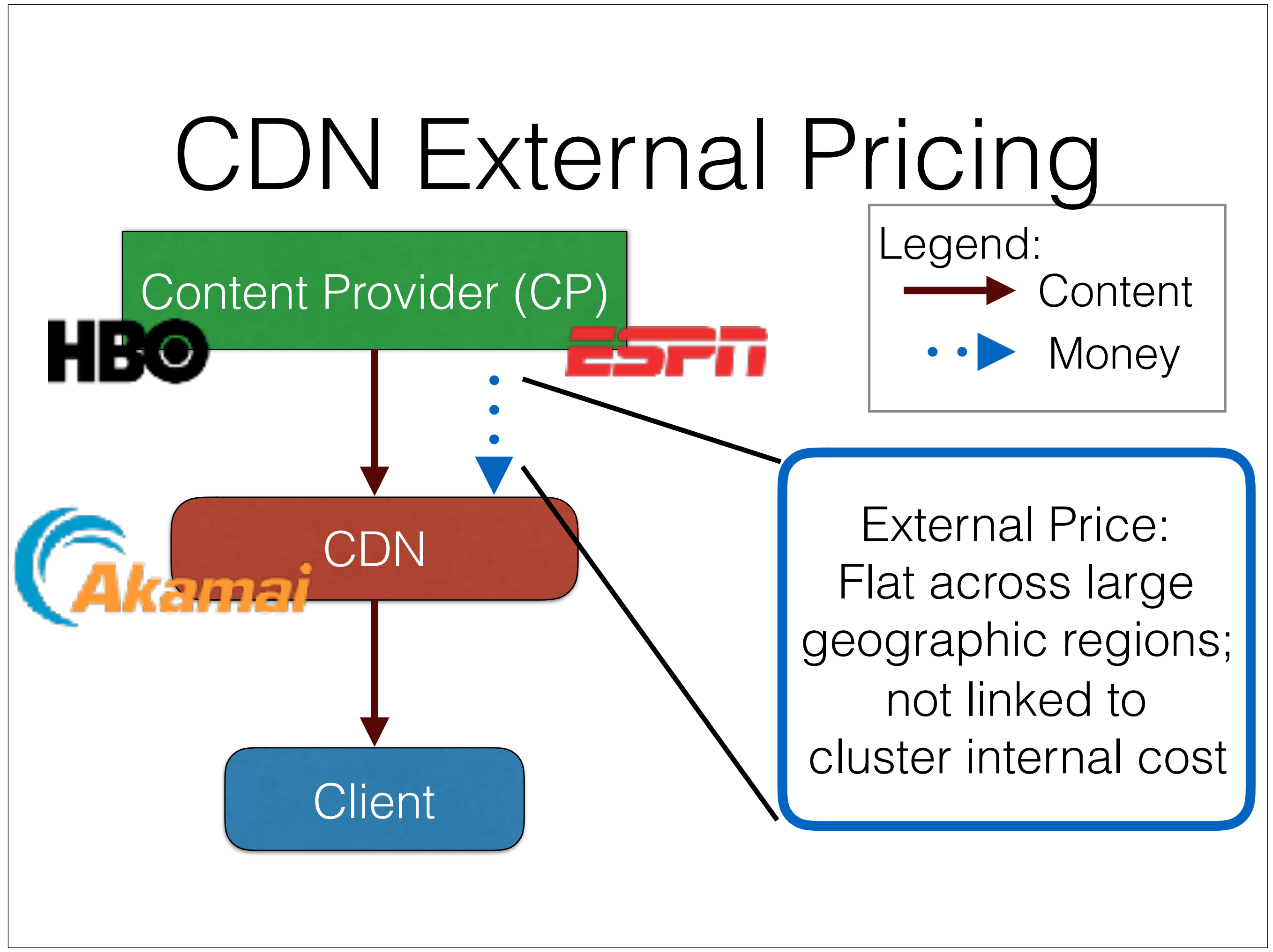


Here we see a zoomed in map of Europe. We have ** many clients in Poland and just one in Germany, but more importantly, we have ** ** two CDNs, one with multiple clusters. We’re going to represent the delivery cost for individual CDN clusters as dollar signs listed on each cluster.

CDN Internal Cost

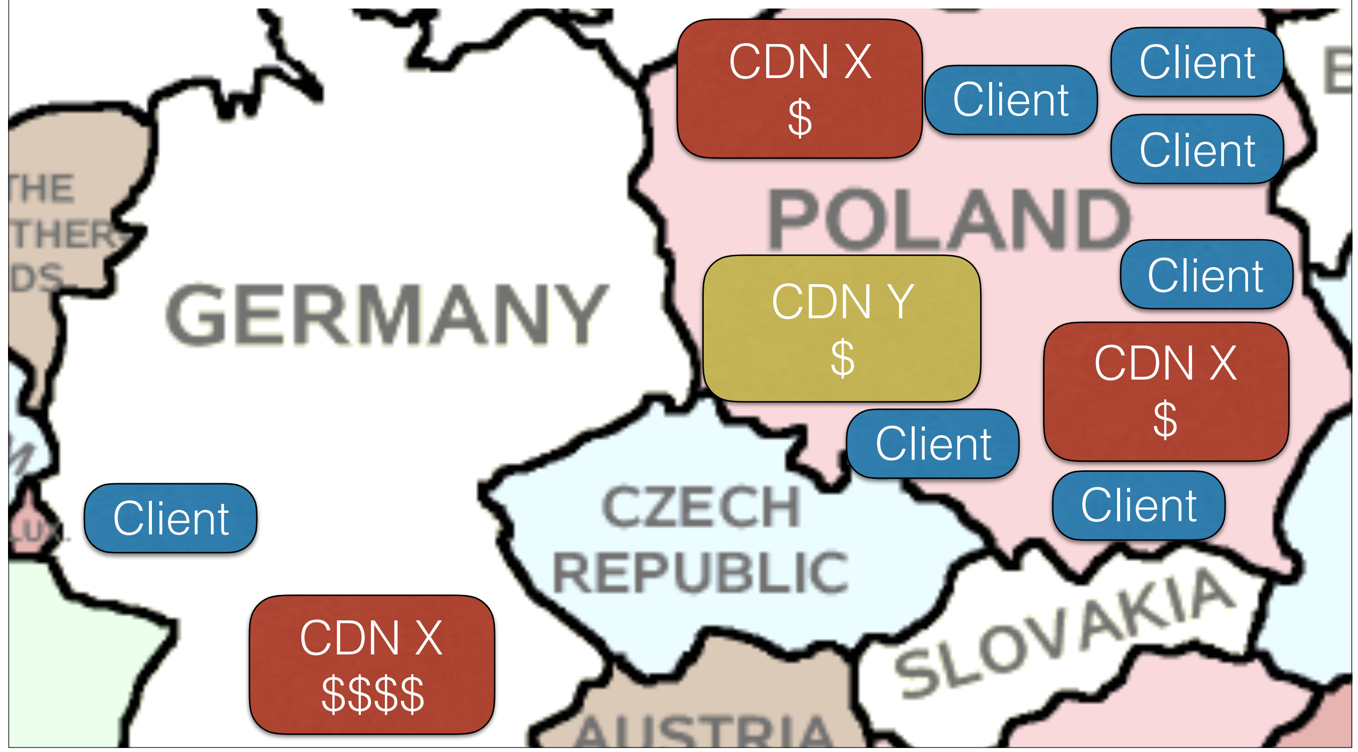


So here we see all the clusters in Poland are cheap to deliver from, but the cluster in Germany is very expensive. With that picture in mind, let’s know learn how CDNs price their services.



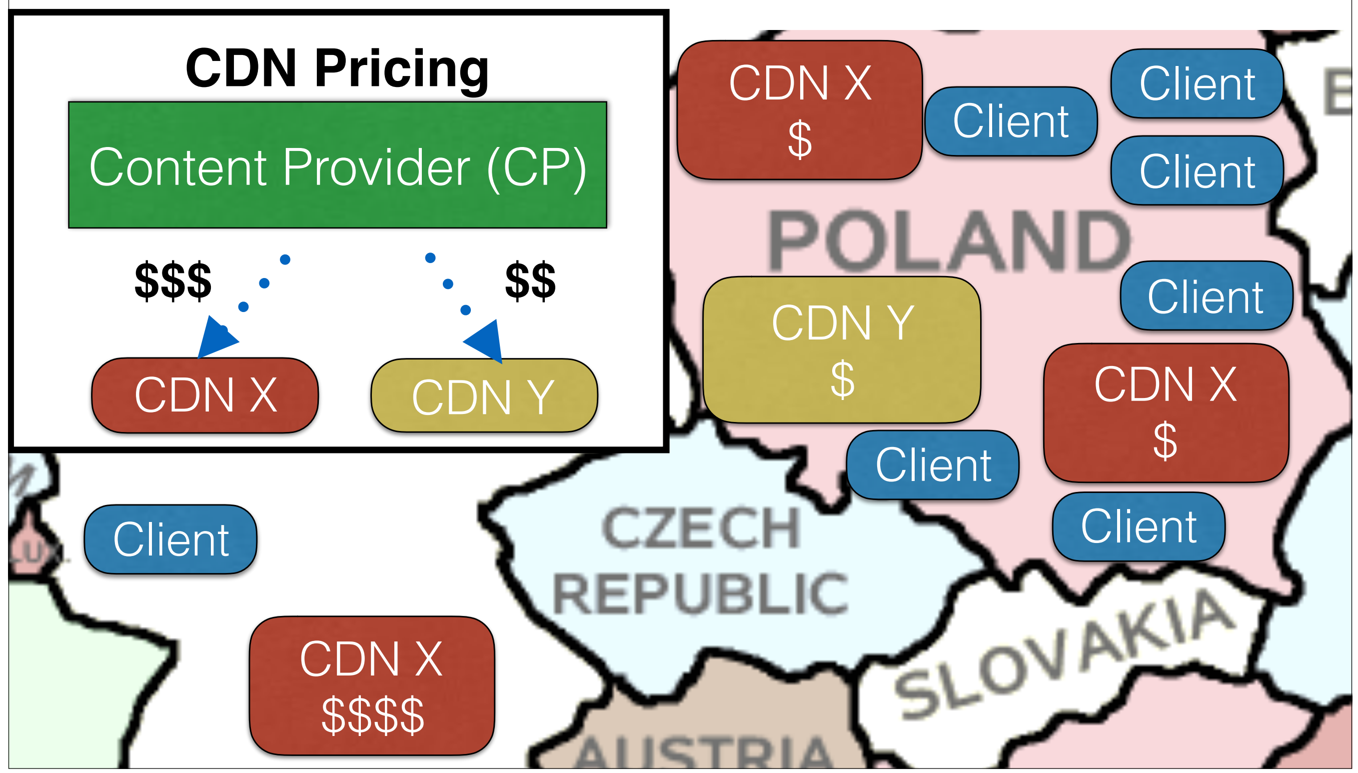
Content providers ** pay CDNs for their delivery services. CDNs negotiate their prices with content providers via long-term contracts, ** which generally have fixed prices across large geographic regions (e.g., continents). The key point is that these prices don't relate to specific clusters' delivery cost.

CDN External Price

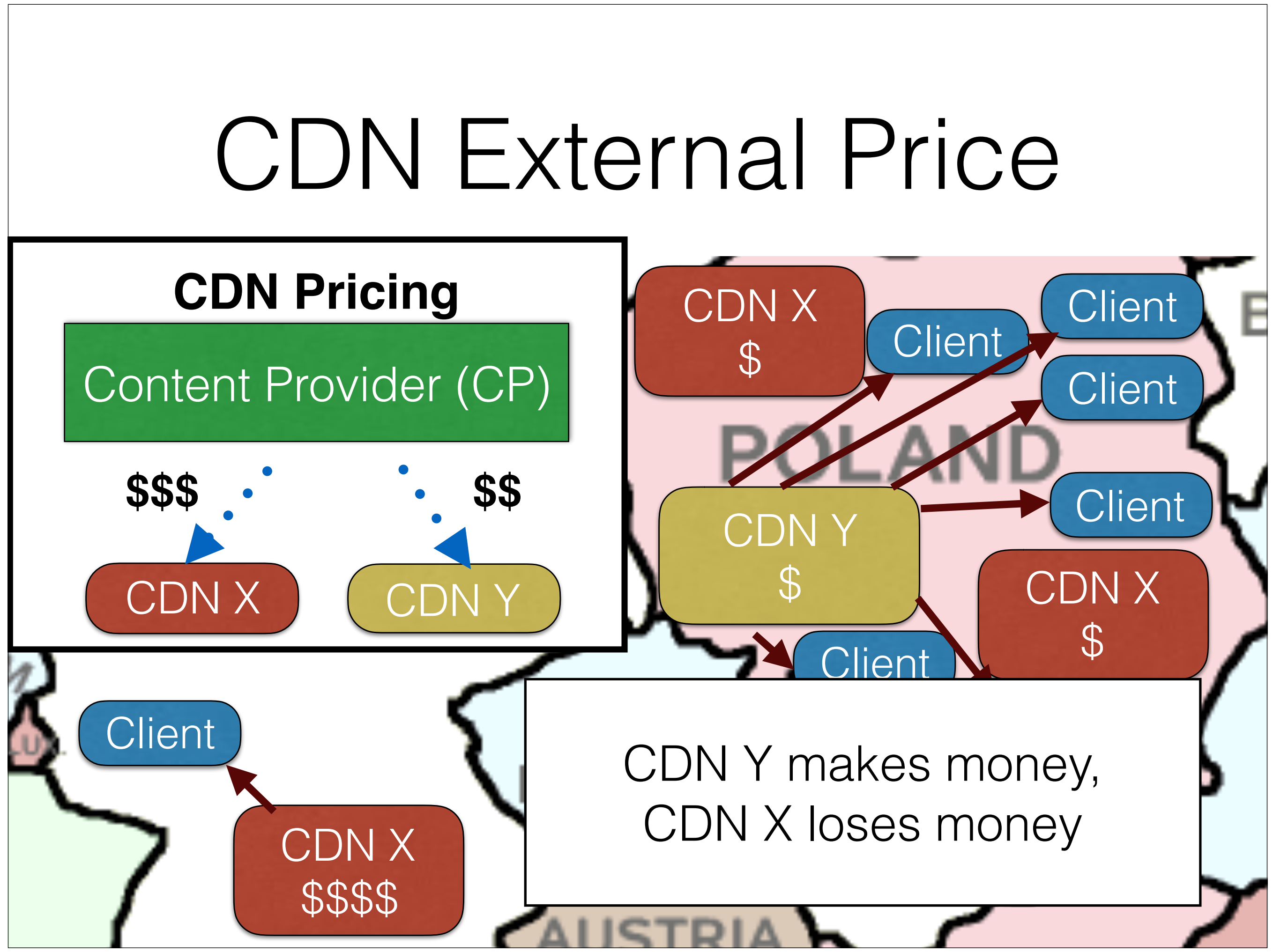


Going back to our hypothetical example; let's say …

CDN External Price

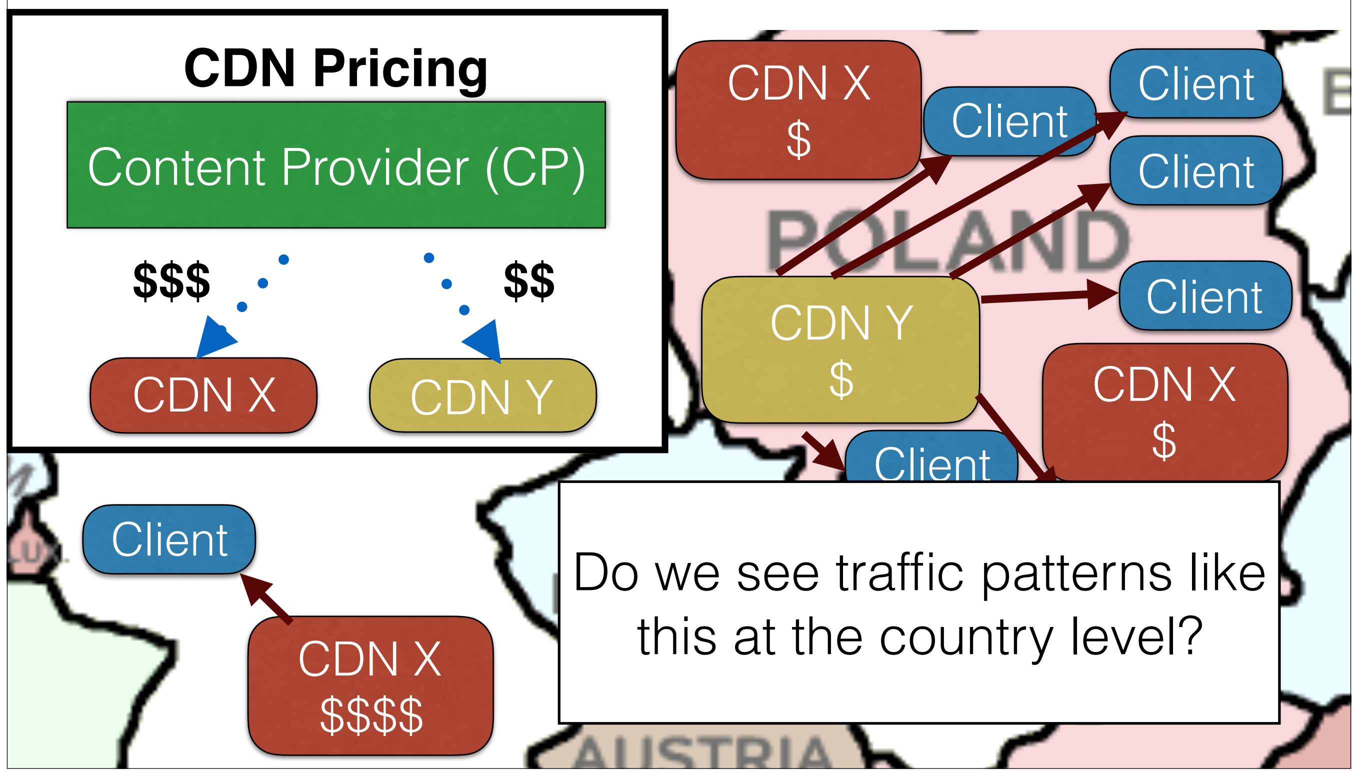


… the content provider negotiated the following contracts with CDN X and Y: it will pay X three dollar signs per byte delivered and Y two dollar signs per byte delivered. As I mentioned before, these are flat rates across whole continents. If that’s the case, a broker might allocate clients …



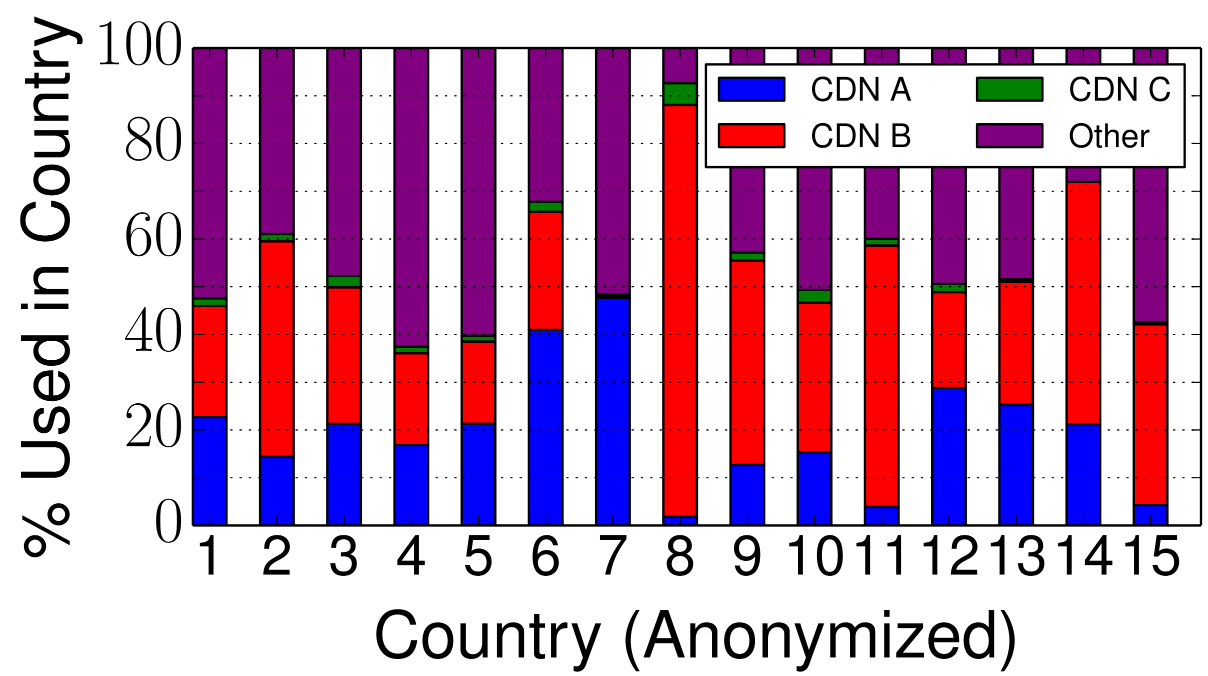
… like this. All clients in Poland go to CDN Y as it's cheaper and can provide adequate performance, and the client in Germany goes to CDN X as it is the only option that can provide adequate performance. Clearly, ** CDN Y makes money as its spends one dollar sign on delivery, yet charges the content provider two dollar signs. However, CDN X loses money as it is charging three dollar signs to the content provider, but only delivers data from it's expensive four dollar sign German cluster. If some of its cheaper Poland clusters were used it could make money, but they are avoided in favor of the cheaper CDN Y.

CDN External Price

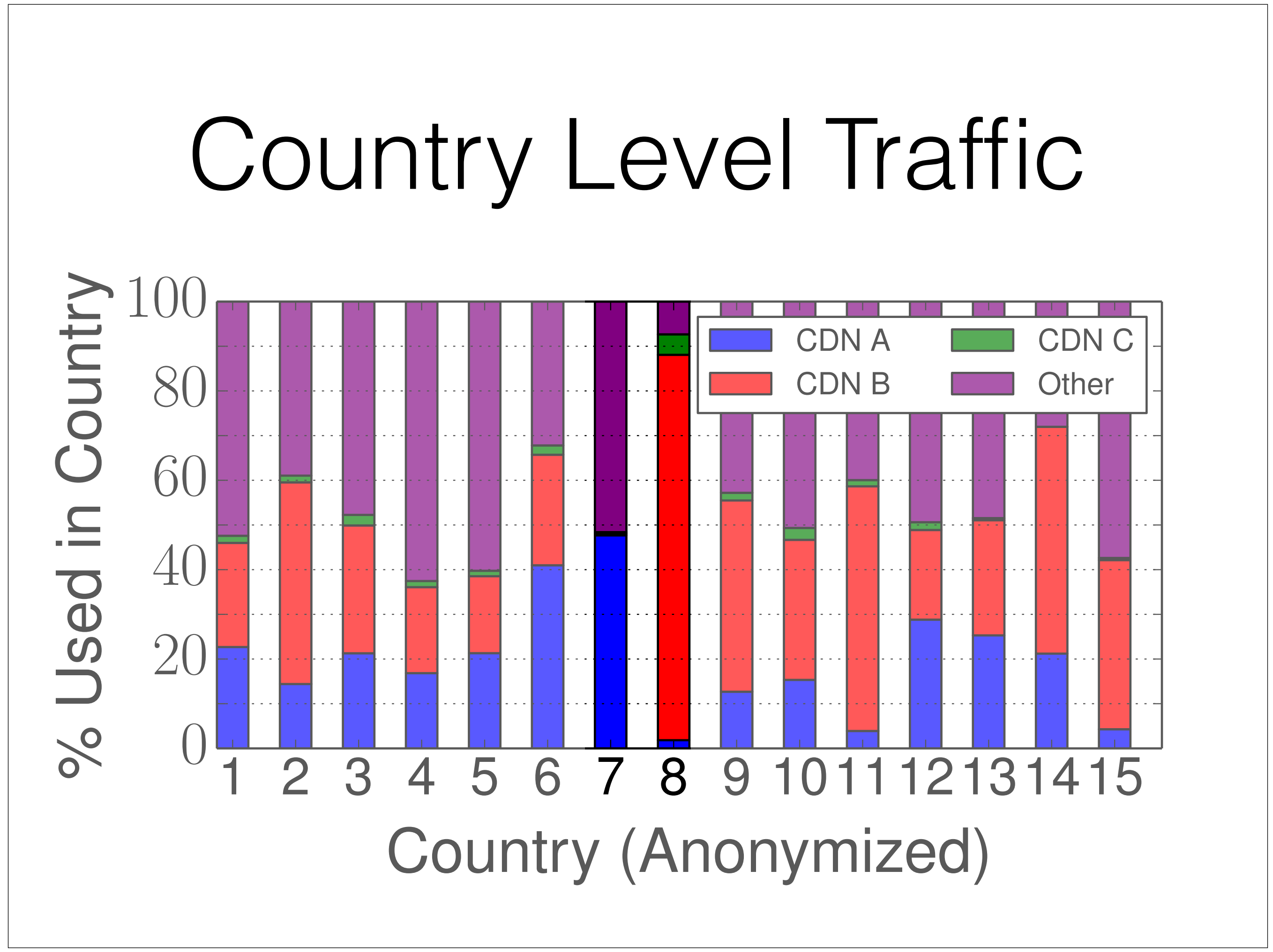


We want to know if something like this actually happens, so let's look at data from a broker to see if traffic patterns like this exist at the country level.

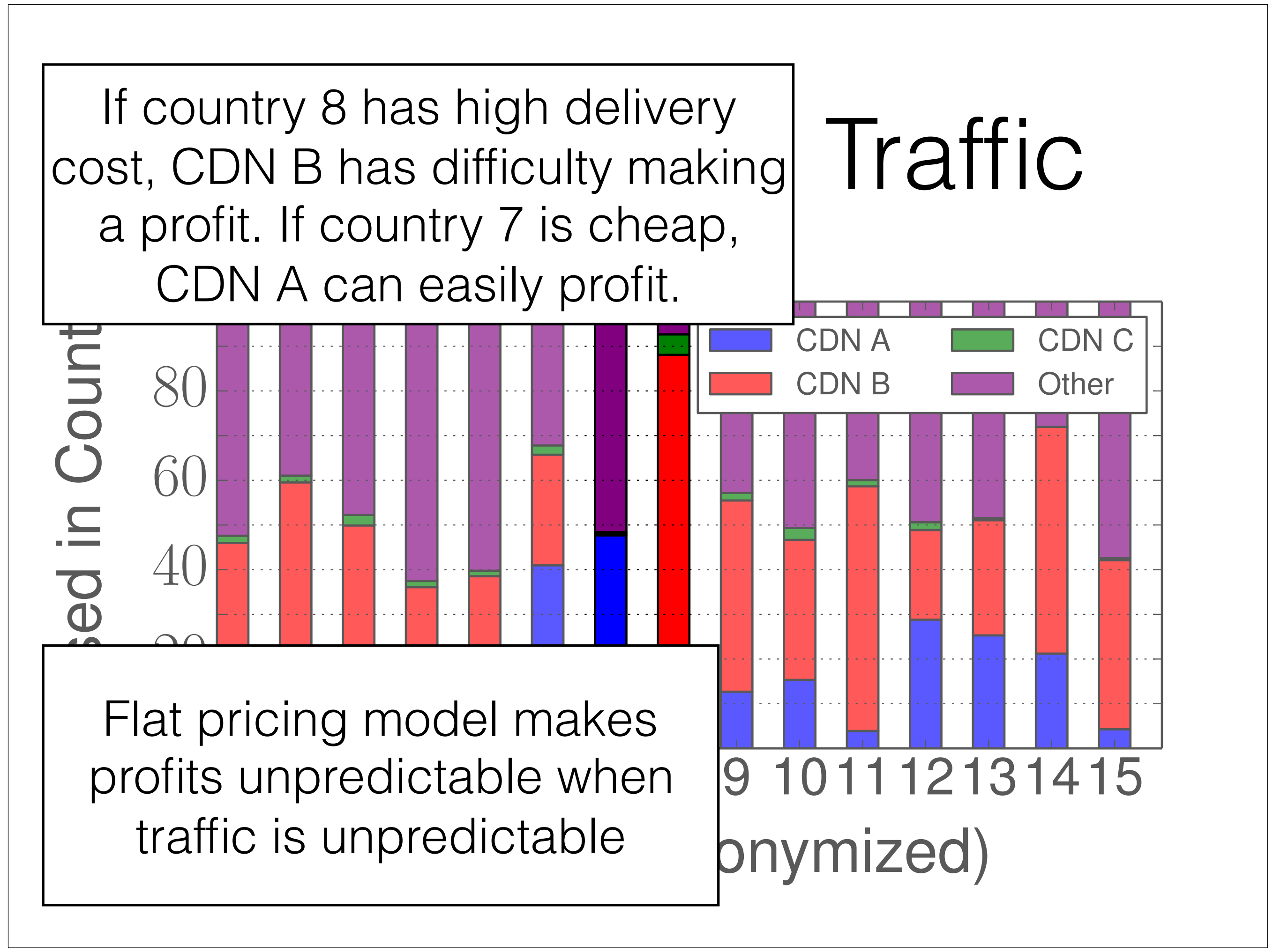
Country Level Traffic



Here we see broker data with client requests binned by country. On the x-axis we show the 15 countries with the most requests. The y-axis shows which CDNs served what percentage of clients in each country. I want to point out two key points of interest: …



… country 8 is predominantly served by CDN B, with few clients served by CDN A. Country 7, on the other hand, is served mostly by CDN A, with few clients served by CDN B. Recall that there's a 30 times variation in cost between some countries.



With that in mind, if country 8 is one with high delivery cost, CDN B has difficulty making a profit. If country 7 has low delivery cost, CDN A can easily profit. What this all points to is the larger problem, ** the CDN flat pricing model makes profits unpredictable when traffic is unpredictable (e.g., due to brokers).

Potential Problems

CDN

- Brokers cause CDN traffic to be unpredictable at short and long timescales making provisioning difficult
- When traffic is unpredictable, flat pricing model makes profits unpredictable

Broker

- Coarse CDN-level selection complicate meeting CP goals
- Incomplete measurements complicate meeting CP goals
- Difficult debugging

See Paper

Now that we understand some problems facing CDNs in a world with brokers, let's talk about how we can fix these problems.

Potential Problems

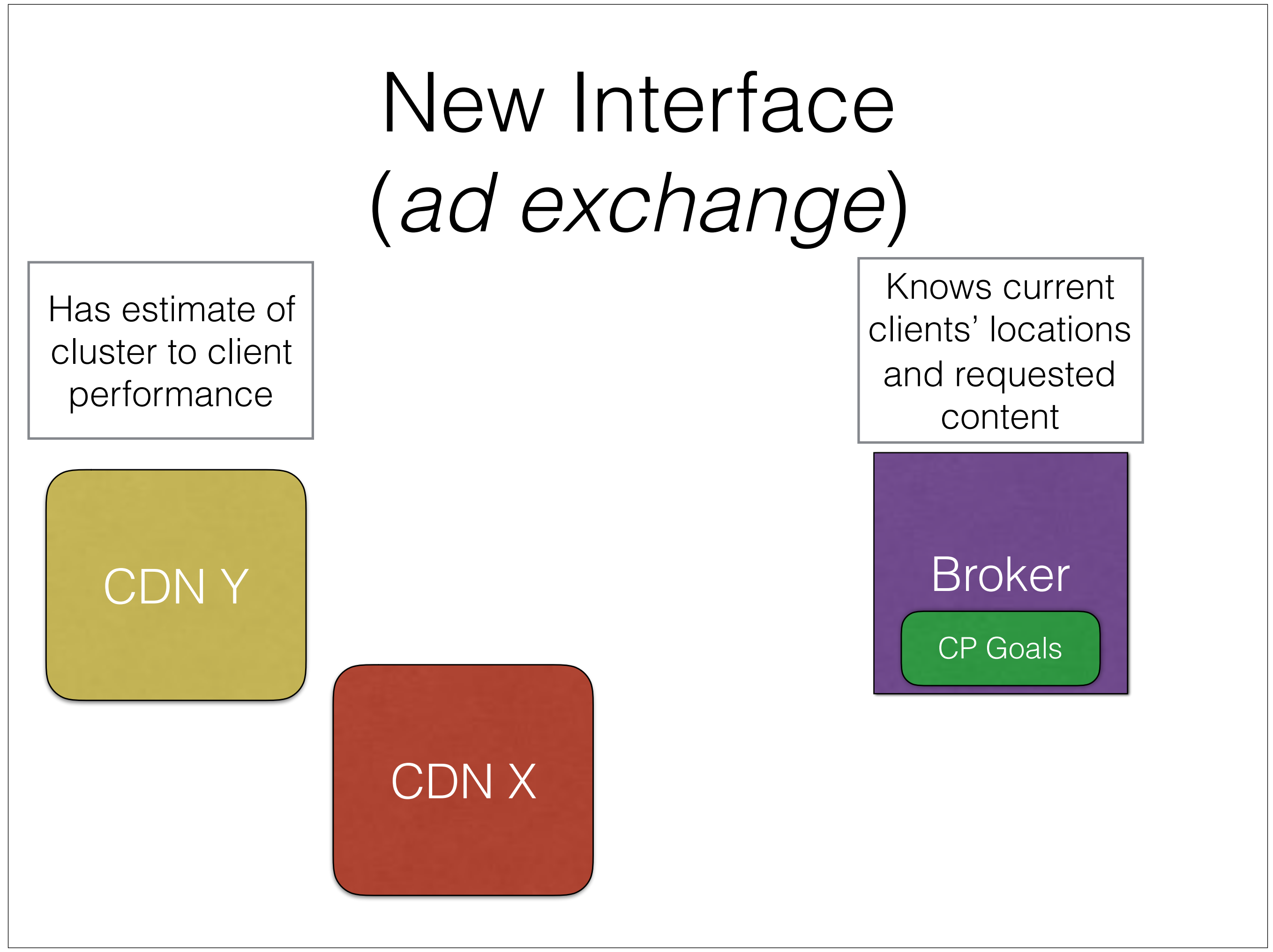
CDN

- Brokers cause CDN traffic to be unpredictable at short and long timescales making provisioning difficult
- When traffic is unpredictable, flat pricing model makes profits unpredictable

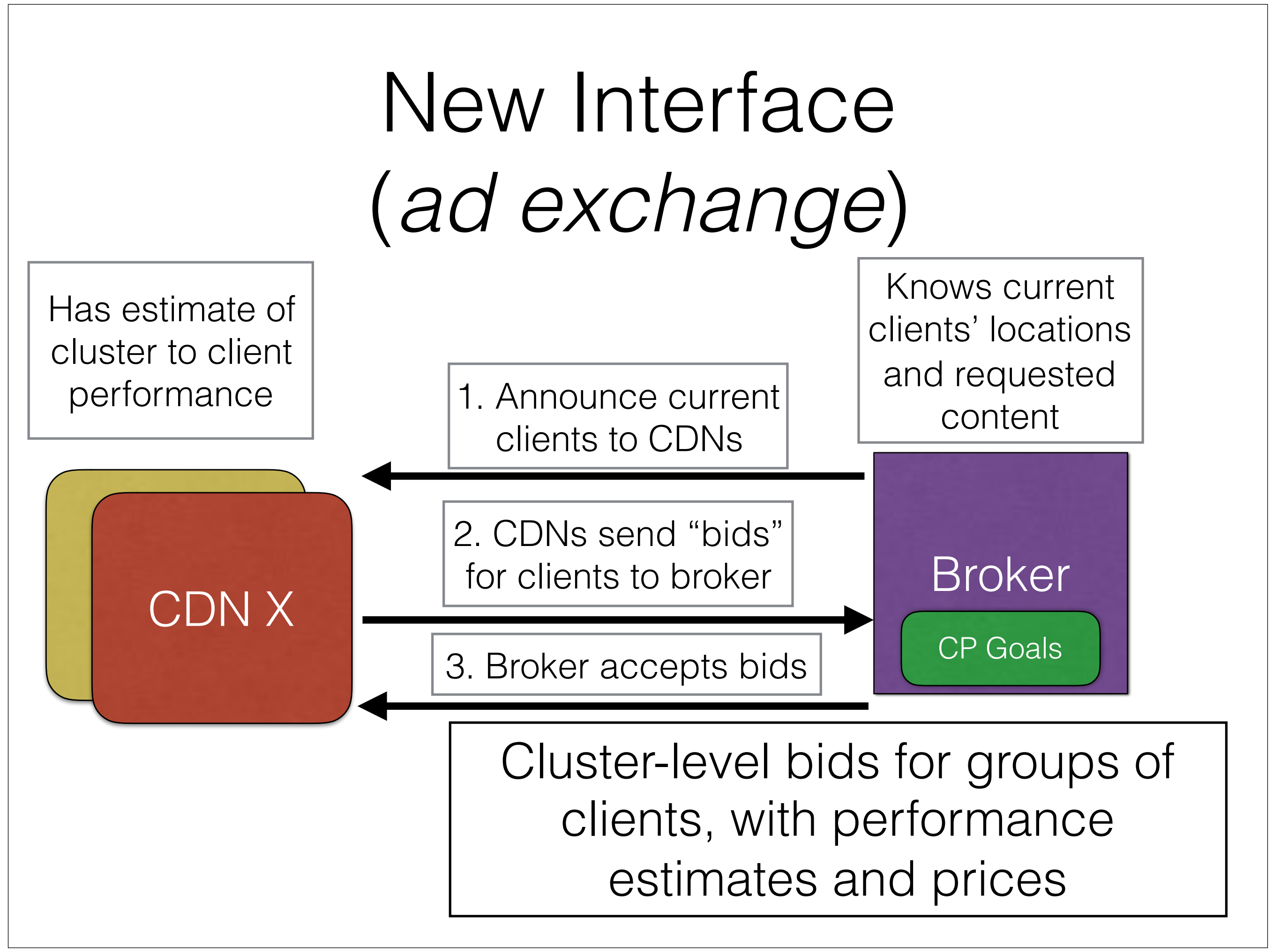
→ Make client allocation explicit

→ Pricing should reflect cost

We can address unpredictable traffic by ** making client allocation explicit among CDNs, and address the flat pricing model problem by ** have pricing reflect cluster delivery cost. We propose a CDN-broker interface that directly addresses these problems using these two ideas, taking inspiration from…

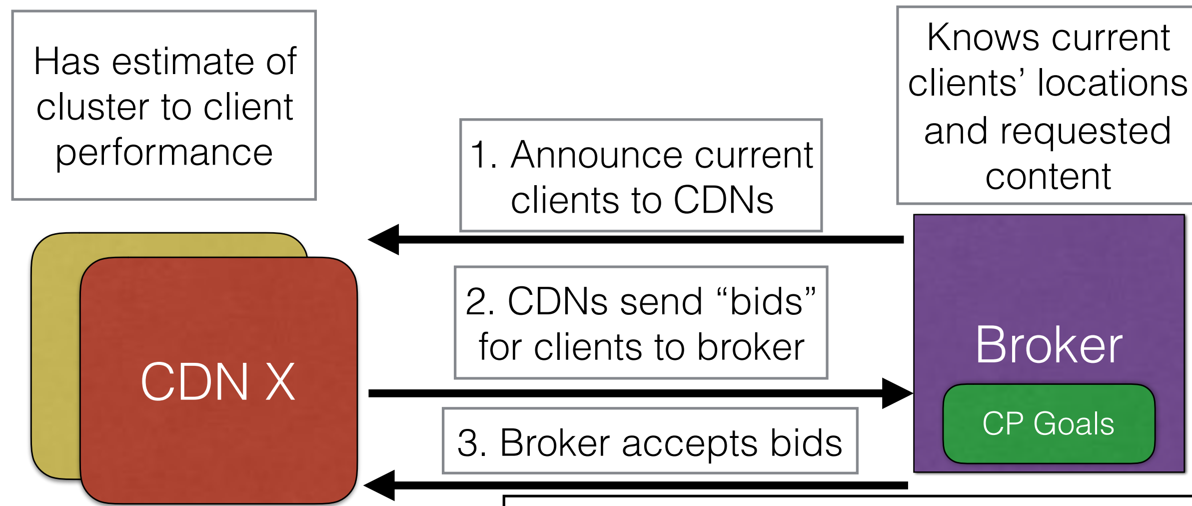


… an ad exchange. As with before, ** CDNs still estimate cluster to client performance. Brokers still ** know about current clients' locations and what content they've requested. But, from here things differ.



Our CDN-broker interface is a control plane protocol that runs in the background periodically. Conceptually, it consists of three stages, here drawn as arrows. First, ** the broker announces the current set of clients to all CDNs. Second, ** the CDNs send "bids" for clients to the broker. These bids ** are done per cluster for groups of clients, with performance estimates and some notion of price. Finally, ** the broker sends back a list of accepted bids to the CDNs.

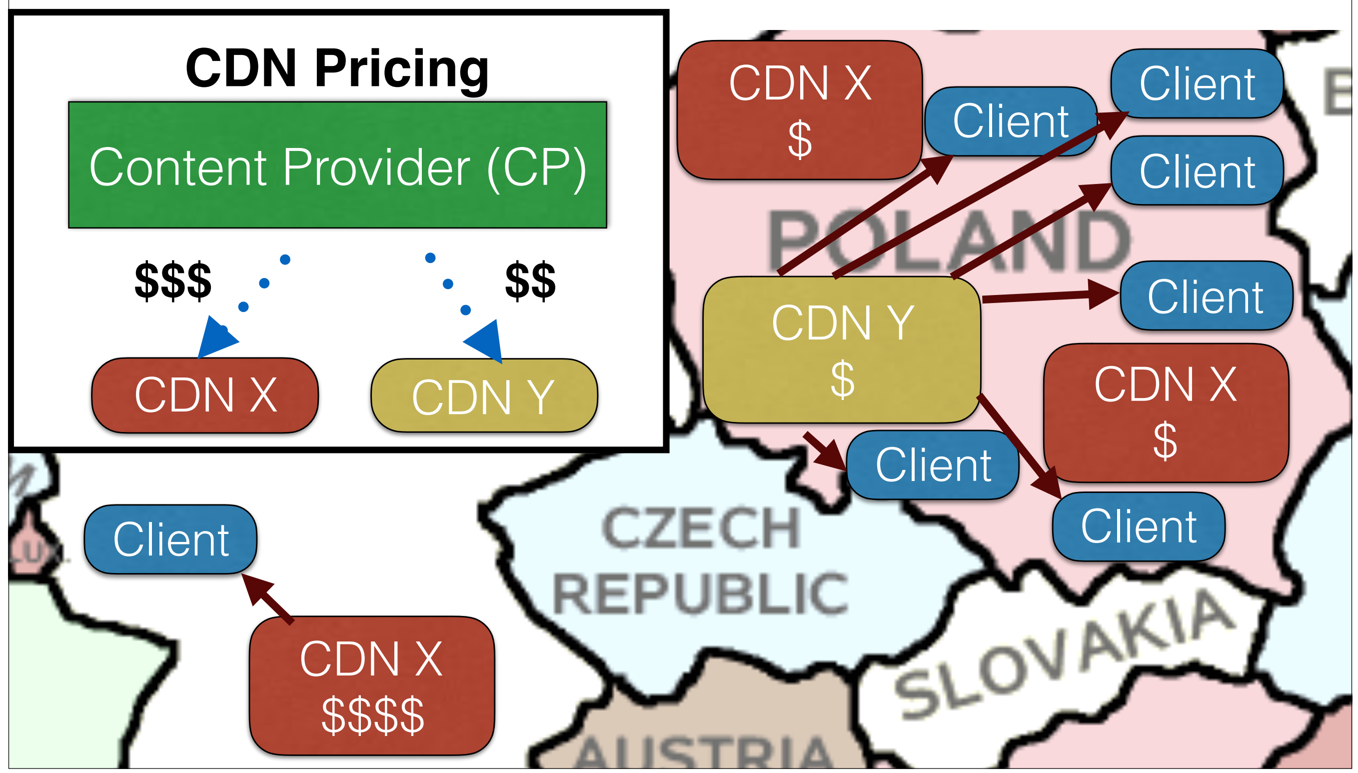
New Interface
(*ad exchange*)



Does this interface address our previous problems?

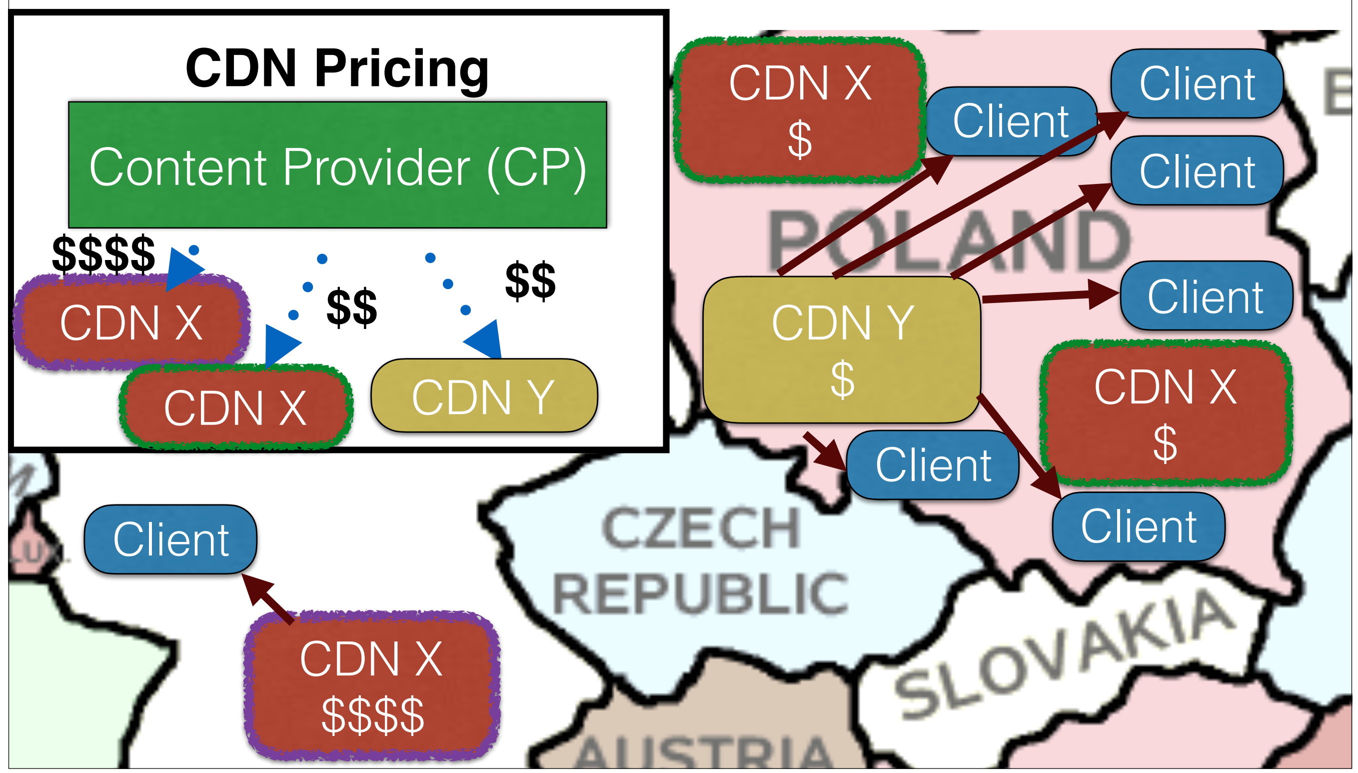
Let's look at how this new proposal addresses the problems we saw previously.

Example



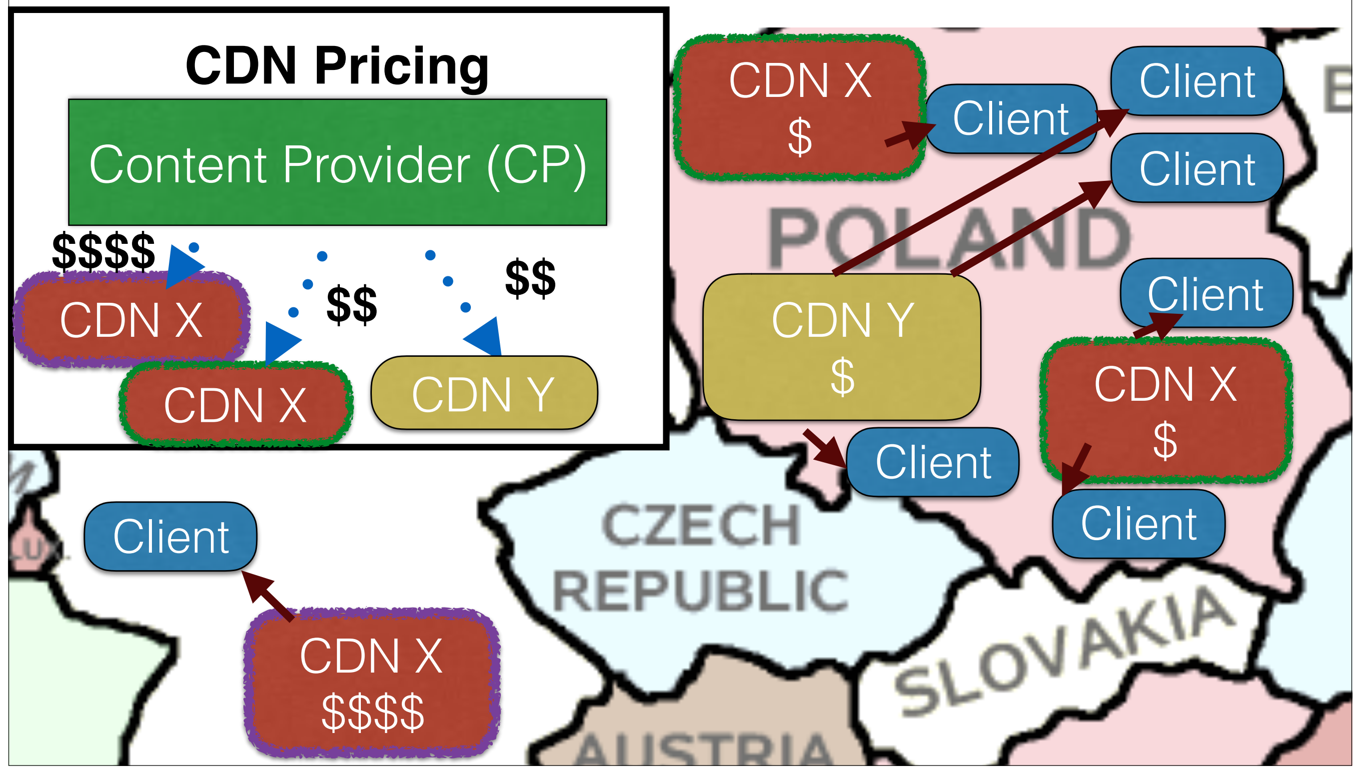
Recall this example. The problem was that CDN X is making less money than it's spending, as it's expensive German cluster is the only one used by the broker. With our proposal, individual clusters within the CDN can have bids with different prices reflecting their delivery cost. Let's fix this example by splitting CDN X into two different groups, Germany and Poland.

Example



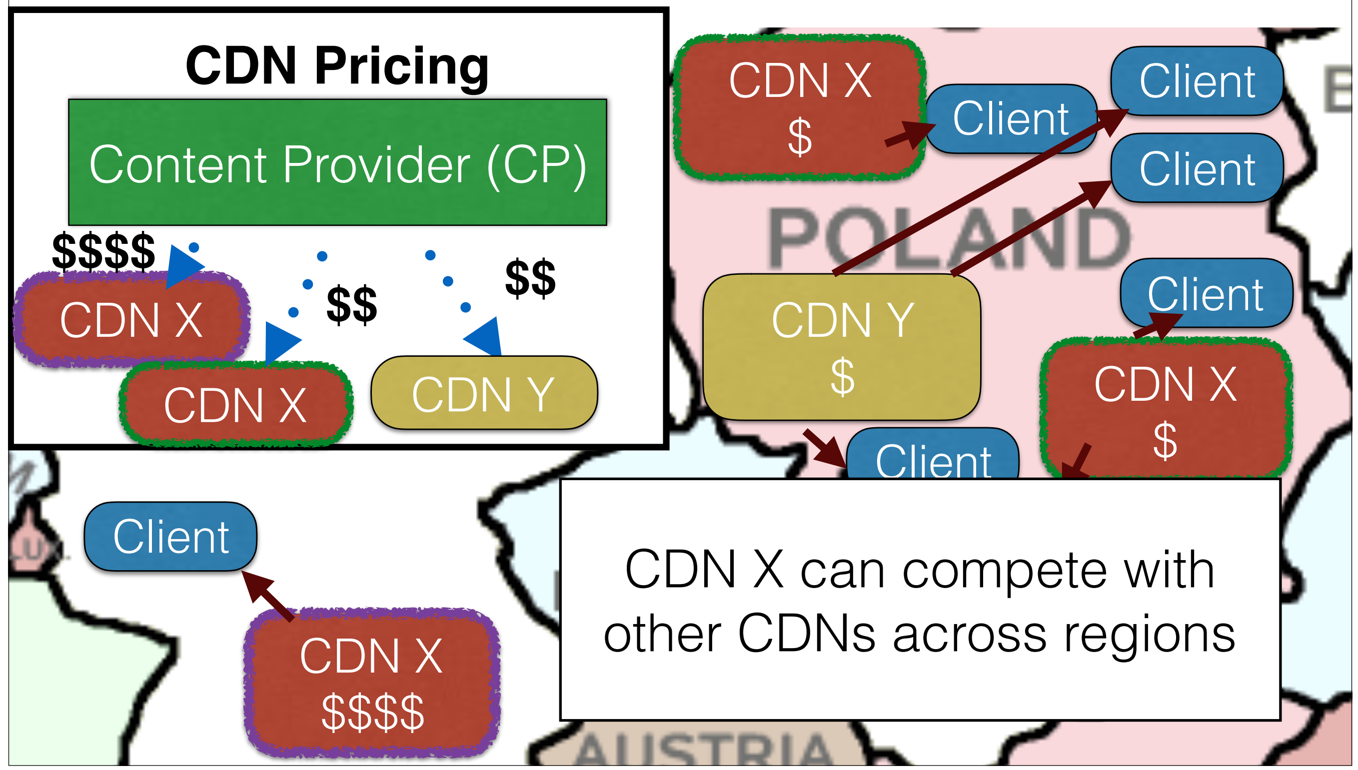
Now, CDN X can price their German cluster at cost, while pricing their Poland clusters to be competitive with CDN Y. As CDN X is now priced competitively in Poland, the broker …

Example



… may move some traffic in Poland to CDN X, …

Example



… allowing CDN X to compete with other CDNs across regions.

Potential Problems

CDN

- Brokers cause CDN traffic to be unpredictable at short and long timescales making provisioning difficult
- When traffic is unpredictable, flat pricing model makes profits unpredictable

Broker

- Coarse CDN-level selection complicate meeting CP goals
- Incomplete measurement complicate meeting CP goals
- Difficult debugging

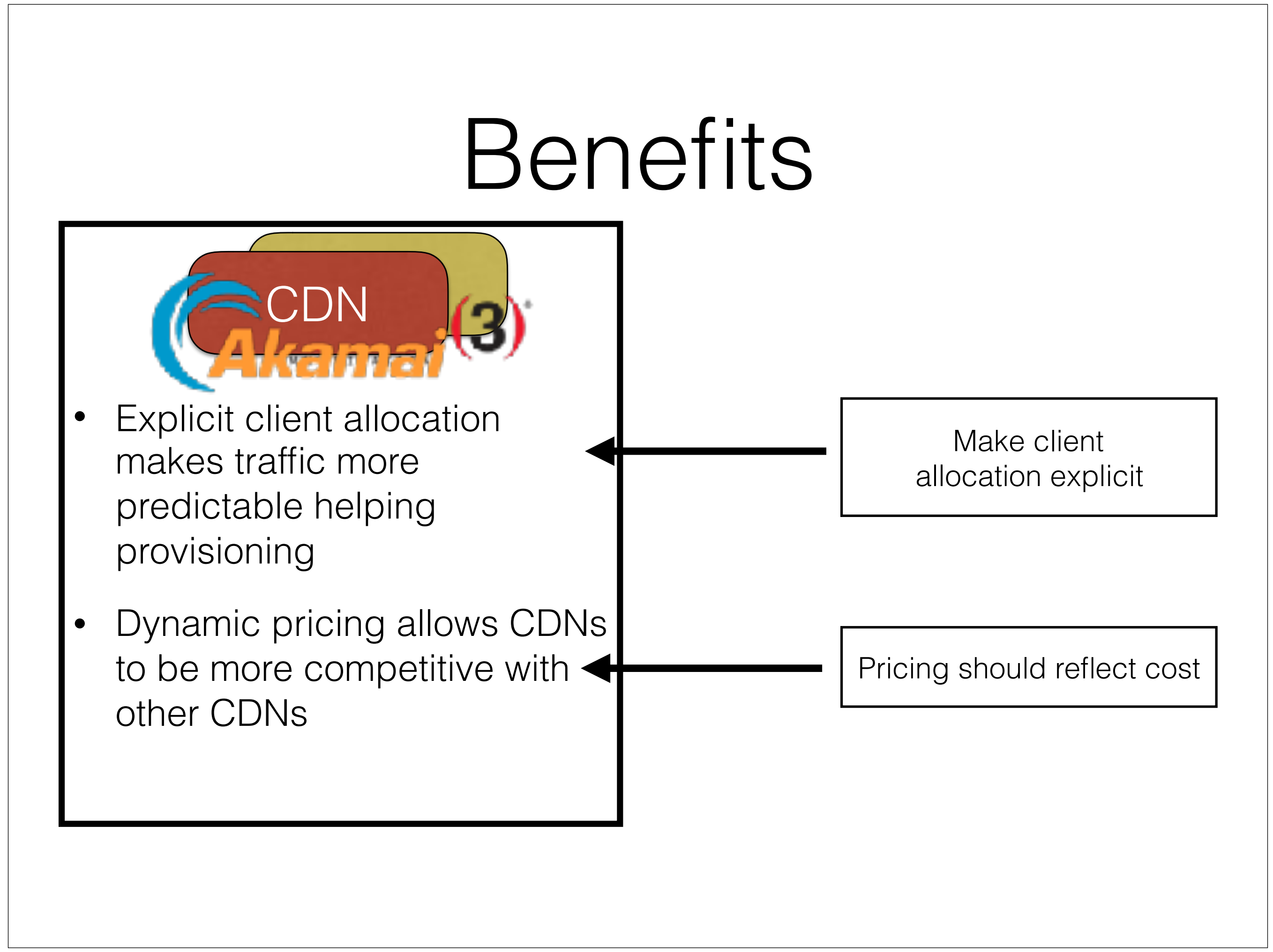
CDNs and brokers clearly can cause issues for one another. ** Some of these issues we examined in depth, which …

Potential Problems

CDN

- Brokers cause CDN traffic to be unpredictable at short and long timescales making provisioning difficult → Make client allocation explicit
- When traffic is unpredictable, flat pricing model makes profits unpredictable → Pricing should reflect cost

… led to some natural solutions. Using those solutions, we propose a CDN-broker interface …



… that can help CDNs make traffic more predictable to ease provisioning, as well as allow CDNs to be more competitive, potentially leading to increased profits.

Déjà Vu

Goal:
ISP stitching

Enabler:
BGP

Result:
10+ years of research

Goal:
CDN stitching

Enabler:
Brokers

Result:
Call to arms!

Finally, I want to draw an analogy between CDN stitching and ISP stitching at the dawn of the Internet. Both have an enabler: BGP or brokers. ISP stitching though spawned 10+ years of research, yet as a community, we've just scratched the surface of the problems in CDN stitching. There's interesting research still to be done here. With that, I'm happy to take questions. Thanks!

Questions from Audience

- Why would this work when CDN federation has not?
  - Broker & CDNs both serve CP. It's a tweak to an existing market, not creation of a new one.
- Does the bidding protocol make actual content delivery slower?
  - No, the bidding protocol is a periodic offline control plane protocol. The data plane is still the same as today.
- Why "auctions"? Why not dynamic pricing?
  - Dynamic pricing can be just as difficult (e.g., stability, convergence, fairness), rearchitecting might be the best approach. Also, there are other gains we didn't touch on in the talk (e.g., ability to use CDN clusters that the broker can't currently see— see paper)