Piracy Statistics: Building on a Rocky Foundation

Published date: 2020-07-23

Alan Ogilvie | Lead Product Manager, Friend MTS

"There are three kinds of lies: lies, damned lies, and statistics" – Unknown

Lies and Piracy Statistics, what is at stake?

The problem of premium content piracy has garnered a lot of attention in the press during the pandemic. With more live sports coming back to screens and action hungry fans unable to attend in-person events or large gatherings, it would make sense that live sports piracy, and all premium content piracy, is on the rise – but how can we tell and what is at stake?

Yes, legitimate content owners and providers have a hard time competing with pirates aggregating content on a large scale due to restrictions of complex and expensive licensing agreements. In the current uncertain economic climate, shady content sources are likely to be increasingly attractive to consumers as seemingly cheaper, albeit less trustworthy and reliable, options. Tracking and measuring the extent of piracy is absolutely necessary to any content protection plan, but if that plan is based on fabricated data – which we see released to the public all too often – creativity, content and revenue growth suffer. **Reliable, Responsible Reporting** is what content owners and distributors need to inform and drive a steady foundation for their strategic content protection decisions.

A Rocky Foundation

How reliable is the piracy data we see published? Various articles highlighting the problem of video piracy are generously strewn with figures aiming to demonstrate the scale of the problem. Compiled using questionable sources and methods, and taken completely out of context, these figures don't hold value under scrutiny, but make headlines. Spreading like clickbait wildfire across various publications and even cited in some studies, this data represents a certain threat of misinformation.

If financial decisions made by investors and licensors are based on inflated and unchecked data, there will be adverse effects on both premium content production and licensing. Consider the following examples of the impact:

- Investors might think twice before making a positive decision on providing funds to filmmakers where perceived certainty about recouping their investment and extent of profit could be impacted.

Friend MTS

- Licensors of premium/exclusive content (e.g. a premium sports Pay-Per-View event) could see the costs of licences increase with perceived rampant content theft.

- Distribution revenues made on a subscription or transactional basis (e.g. a monthly subscription to a legitimate OTT service or a virtual ticket for a Pay-Per-View event) are impacted as businesses try to battle the apparent "massive" piracy for paying customers – driving down cost expectations of legitimate services by trying to compete with the pirate services.

So, while we know that piracy is not a victimless crime, skewed piracy statistics are not harmless either. They have a detrimental impact on investment and distribution revenues.

What *Lies* in Video Piracy Data – Supply & Demand

What do these numbers actually represent? What kind of data are they based on? And why do we consider these numbers inaccurate?

Some of these numbers strive to represent supply of pirated content, while others the consumer demand for it. However, neither seems to be actually represented.

The **supply** of illegitimate content is presented in many reports with a number of URLs/links. While peripherally representing supply of content on websites promoting piracy (largely ignoring non-web based piracy distribution channels along the way), this approach leads to inaccurate data due to the fact that many of these links lead to websites aggregating piracy links, rather than to websites actually hosting illegitimate content.

At best, the data represented provides some insight into the number of hyperlinks to illegitimate web-based content. It provides no data about its **supply**, nor consumption patterns across **all** piracy sources or devices.

As for the figures on **demand** for illegitimate content, there are at least two fundamental issues:

Issue 1: Figures do not represent consumption as they are not based on direct measurement

For *legitimate* content streamed over the internet, for example, video content consumption data can come from servers or network providers. In some regions, pay-TV and streaming services have to provide statistics to a reputable reporting organisation that aggregates data for analysing various viewing trends, e.g.: Nielsen in the U.S. and Ofcom in the UK.

Therefore, even for legitimate content, unless the data on video consumption comes from any or all of the following entities the data would be unreliable.

- The server operations team from which the streams originate (stream playback analytics data).

- The network providers that it transmits across; however it is difficult to ascertain without data on which specific traffic was streaming illegitimate audio/video material.

- One of the reputable reporting organisations.

Without reliable data, one cannot establish accurate numbers on how many people are actually consuming the content.

When it comes to *illegitimate* content provided by pirate services, measuring such illegal online activities is problematic even if we assume access to the data from servers, networks or reporting organisations. Due to the furtive nature of piracy, companies and organisations that would like to present generic 'piracy statistics' would find this impossible without close cooperation with other industry players such as ISPs who can inspect their traffic and understand what is piracy and what isn't. This level of data has never been published openly for scrutiny, verification and reporting.

- Some organisations providing 'piracy statistics' start with a base of links where text string patterns match prominent content titles (feature films, albums, etc.). Then they try to extrapolate how many people may have seen the links using volume data from non-piracy consumer web trends and position this to present a worst-case scenario for piracy. Not only is this 'rocky base' built on inaccurate or incomplete data, but the application of data to perform an extrapolation to show how many may be using the pirated services compounds the inaccuracy.

- Some companies that provide the data to allow extrapolation are using analytics data from legitimate websites. This is typically used to allow some forecasting of consumption trends over time, a common scenario for marketing teams when comparing similar website genres & categories. It is important to realise that this extrapolation data source cannot provide insight to visitors on pirate services, which has a markedly different traffic profile to standard non-piracy consumer trends. The pirate sites have no need or desire to provide direct measurement data.

Organisations reporting data created in this way are providing seriously flawed data as it is not based on direct measurement. It is definitely not recommended to equate the number of publicly available links (of possible pirate material) with actual consumption. Of course, the need for publicly available links in their methodology is a problem. We know that private and restricted group chats and forums are often sources of links to pirate material, and not counting these compounds the statistical inaccuracies.

Issue 2: Figures are based solely on web-based piracy

Pulling from our extensive experience and knowledge of the *complete pirate ecosystem*, web-based piracy represents a small minority of consumption compared to smart device apps; streaming into Kodi add-ins, Plex plugins, other software frameworks; streaming set-top boxes (MAG and the like); and other illicit streaming devices (ISDs).

Given the nature of how these 'piracy statistics' reporting organisations gather their base data and extrapolate for reporting, their accuracy is obviously broken since it misses the remaining majority of illicit and illegal online activities. Whether through programmatic means or polling, it is unlikely that those individuals engaged in such activities would volunteer to be monitored or answer surveys honestly.

As a result, one has to question the relevance and accuracy of such statistics. Even taking them as a guide by focusing solely on web-based piracy, they ignore the largest elements of pirate content consumption.

A Solid Foundation

At Friend MTS, we believe that **Reliable, Responsible Reporting** should form a solid foundation for management decisions and such data must:

- Originate from **sources that are statistically more relevant** – looking at the sources of piracy and measuring piracy consumption across network traffic rather than publicly shared links to websites.

- Be representative of **the entire population** it is seeking to depict, not just websites but also streaming applications/OS (Kodi, et. al.), streaming STBs, and other ISDs.

- Be analysed in context to facilitate robust piracy source takedown protecting premium content from theft.

After all, if you build on a rocky foundation of statistically inaccurate numbers, your content protection strategy might not prove to be robust or able to withstand the test of the real-world piracy problem.

If you would like to understand how big the piracy problem really is for your business, contact us today.