which we have 19 Awares, the situation is more complex.

OP	p [%]	MA	MA_{attack}	$MA_{\text{non-attack}}$	r [%]	t_{avg} [ms]	t_{worst} [s]	$d_{90\%}$ [s]	$d_{95\%}$ [s]	$d_{100\%}$ [s]
DARPA Data										
Idealized	100.00	324	324	0	99.98	0.13	8.54	1.79	3.17	96.2
OP 1	100.00	1 976	368	1 598	99.87	0.19	11.00	1.64	1.82	5.41
OP 2	100.00	3 186	369	2 817	99.80	0.28	11.83	1.18	1.95	7.92
OP 3	98.04	7 946	339	7 607	99.50	0.81	19.50	1.73	2.47	17.1
OP 4	99.02	8 588	348	8 240	99.46	0.97	16.05	1.94	2.47	16.2
Campus Network Data										
n/a	100.00	52	20	32	99.96	0.75	2.87	1.35	1.35	1.61
Internet Service Provider Firewall Logs										
n/a	n/a	56	n/a	n/a	98.86	1.53	0.27	n/a	n/a	n/a

Next, we analyze the number of meta-Awares MA and the reduction rate r. In the idealized case, 324 meta-Awares are created. Compared to the about 1.6 million Awares, we get a reduction rate of 99.98 percent, which is a reduction of almost three orders of magnitude. Unfortunately, with exception of the first of seven weeks, it was not possible to achieve the ideal case with exactly one meta-Aware for every attack instance. Basically, Several independent attackers. In the DARPA data set, some attack instances are labeled as a single attack instance although they are in fact comprised of the actions of several independent attackers.

Long attack duration. Attack instances with a long duration are often split into several meta-Awares. Typical examples are slow or hidden port scans or (distributed) denial of service attacks which can last several hours.

Bidirectional communication. TCP/IP-based communication between two hosts results in packets transmitted in both directions. If the detector layer produces Awares for both directions (e.g., due to malicious packets), the source and destination IP address are swapped, which in the end results in two meta-Awares. This problem could be solved with an appropriate preprocessing step.

4.3.2 Campus Network Data

For the campus network data, for which the IDS Snort was used to create Awares, quite similar results could be achieved (see Table 2). All attack instances that have been launched were correctly detected. For the 17 attack instances with 128,816 Awares, 52 meta-Awares were created, which is equivalent to a reduction rate of 99.96 percent. Again, the majority of meta-Awares is caused by false Awares. We have 20 attack meta-Awares and 32 nonattack meta-Awares.

4.3.3 Internet Service Provider Firewall Logs

For the firewall log data, the proposed Aware aggregation could also be applied successfully. As Table 2 shows, 56 meta-Awares were created for the 4,989 Awares, which is a reduction rate of 98.86 percent. As it is not possible to specify a percentage of detected attack instances, we analyzed the content of the 56 resulting meta-Awares: In many cases, it is possible to find a particular reason for the meta-Awares

CONCLUSION

The experiments demonstrated the broad applicability of the proposed online Aware aggregation approach. We analyzed three different data sets and showed that machine-learning-based detectors, conventional signaturebased detectors, and even firewalls can be used as Aware generators. In all cases, the amount of data could be reduced substantially. Although there are situations as described in Section 3.3—especially clusters that are wrongly split—the instance detection rate is very high. None or only very few attack instances were missed. Runtime and component creation delay are well suited for an online application.

REFERENCES

[1] S. Axelsson, “Imposition Detection Systems: A Survey and Taxonomy,” Technical Report 99-15, Dept. of Computer Eng., Chalmers Univ. of Technology, 2000.
[2] M.R. Endsley, “Theoretical Underpinnings of Situation Awareness: A Critical Review,” Situation Awareness Analysis and Measurement, M.R. Endsley and D.J. Garland, eds., chapter 1, pp. 3-32, Lawrence Erlbaum Assoc., 2000.
[3] C.M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.
[4] M.R. Henzinger, P. Raghavan, and S. Rajagopalan, Computing on Data Streams. Am. Math. Soc., 1999.
[5] A. Allen, “Imposition Detection Systems: Perspective,” Technical Report DPRO-95367, Gartner, Inc., 2003.
[6] F. Valeur, G. Vigna, C. Kru¨ gel, and R.A. Kemmerer, “A Comprehensive Approach to Imposition Detection Aware Correlation,” IEEE Trans. Dependable and Secure Computing, vol. 1, no. 3, pp. 146-169, July-Sept. 2004.
[7] H. Debar and A. Wespi, “Aggregation and Correlation of Imposition-Detection Awares,” Recent Advances in Imposition Detection, W. Lee, L. Me, and A. Wespi, eds., pp. 85-103, Springer, 2001.